

GI-Edition

Lecture Notes in Informatics

Douglas W. Cunningham, Petra Hofstedt,
Klaus Meer, Ingo Schmitt (Hrsg.)

INFORMATIK 2015

28.9.-2.10. 2015
Cottbus

246

Proceedings

D. Cunningham, P. Hofstedt, K. Meer, I. Schmitt (Hrsg.): INFORMATIK 2015



Douglas W. Cunningham, Petra Hofstedt, Klaus Meer,
Ingo Schmitt (Hrsg.)

45. Jahrestagung der Gesellschaft für Informatik
INFORMATIK 2015

Informatik, Energie und Umwelt

28. September – 2. Oktober 2015
in Cottbus, Deutschland

Gesellschaft für Informatik e.V. (GI)

Lecture Notes in Informatics (LNI) - Proceedings
Series of the Gesellschaft für Informatik (GI)

Volume P-246

ISBN 978-3-88579-640-4
ISSN 1617-5468

Volume Editors
Prof. Dr. Douglas W. Cunningham
Prof. Dr. Petra Hofstedt
Prof. Dr. Klaus Meer
Prof. Dr. Ingo Schmitt
Brandenburgische Technische Universität Cottbus-Senftenberg
Institut für Informatik
Platz der Deutschen Einheit 1
D-03046 Cottbus, Germany
E-Mails: {douglas.cunningham,hofstedt,meer,schmitt}@b-tu.de

Series Editorial Board
Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Austria
(Chairman, mayr@ifit.uni-klu.ac.at)
Dieter Fellner, Technische Universität Darmstadt, Germany
Ulrich Flegel, Hochschule für Technik, Stuttgart, Germany
Ulrich Frank, Universität Duisburg-Essen, Germany
Johann-Christoph Freytag, Humboldt-Universität zu Berlin, Germany
Michael Goedicke, Universität Duisburg-Essen, Germany
Ralf Hofestädt, Universität Bielefeld, Germany
Michael Koch, Universität der Bundeswehr München, Germany
Axel Lehmann, Universität der Bundeswehr München, Germany
Peter Sanders, Karlsruher Institut für Technologie (KIT), Germany
Sigrid Schubert, Universität Siegen, Germany
Ingo Timm, Universität Trier, Germany
Karin Vosseberg, Hochschule Bremerhaven, Germany
Maria Wimmer, Universität Koblenz-Landau, Germany

Dissertations
Steffen Hölldobler, Technische Universität Dresden, Germany
Seminars
Reinhard Wilhelm, Universität des Saarlandes, Germany
Thematics
Andreas Oberweis, Karlsruher Institut für Technologie (KIT), Germany

© Gesellschaft für Informatik, Bonn 2015
printed by Köllen Druck+Verlag GmbH, Bonn

Vorwort der Herausgeber

Die 45. Jahrestagung der Gesellschaft für Informatik (GI), die INFORMATIK 2015, fand vom 28. September 2015 bis zum 2. Oktober 2015 auf dem Campus Cottbus der Brandenburgischen Technischen Universität Cottbus-Senftenberg statt. Das Motto der Jahrestagung 2015 lautete "Informatik, Energie und Umwelt". Dementsprechend wurden bevorzugt die Zusammenhänge dieser drei für das gesellschaftliche Leben so zentralen Bereiche diskutiert.

Schwerpunkt dabei waren Fragen wie etwa: Welche Rolle spielt die Informatik bei der Entwicklung und Analyse umweltrelevanter Systeme (Umweltinformatik), welchen Nutzen hat sie für den Entwurf ökologisch nachhaltiger und energieeffizienter(er) Systeme sowie für die bessere Nutzung von Ressourcen, was kann die Informatik zur Optimierung von Energieträgern beitragen? Auch von diesen Schwerpunkten unabhängige Themen wurden in Form von Workshops und Tutorien diskutiert.

Dieser Band enthält rund 160 Beiträge zu den einzelnen Workshops. Die Auswahl der Workshops wurde von der Tagungsleitung vorgenommen, jeder Workshop hatte ein Programmkomitee, das für die Begutachtung und Auswahl der schriftlichen Beiträge verantwortlich war.

Das Tagungsprogramm bestand aus dem Tag der Informatik am 30.9.2015 mit 4 eingeladenen Hauptvorträgen, an den anderen Tagen aus 26 Workshops, 7 Tutorien, Studenten- und Doktorandenprogramm sowie einem Industrieprogramm. Die Durchführung einer Tagung dieser Größe erfordert viel Unterstützung von unterschiedlichster Seite. Wir danken an dieser Stelle allen Sponsoren, den Hauptvortragenden Prof. Dr. Deussen, Dr. Pappe, Prof. Dr. h. c. mult. Plattner und Prof. Dr. Sonnenschein sowie Prof. Dr. Dueck für seinen Vortrag bei der Festveranstaltung, Frau Winter und ihrem Team von der GI sowie den vielen Helferinnen und Helfern vor Ort. Ebenso danken wir allen Teilnehmerinnen und Teilnehmern. Die Jahrestagung stand unter der Schirmherrschaft von der Bundesministerin für Bildung und Forschung Frau Prof. Dr. Wanka

Cottbus, Juli 2015

Douglas W. Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt

Sponsoren

Wir danken den folgenden Unternehmen und Institutionen für die Unterstützung der Konferenz.

Harting – Technology Group – Pushing Performance

Hugendubel Fachinformationen GbmH

IHP – Innovations for High Performance Microelectronics

Infineon Technologies AG

Micromata GmbH

Orga Plan Cottbus Beratungs- und Vertriebsgesellschaft mbH

Philotech Systementwicklung und Software GmbH

PROFI – Engineering Systems AG – Innovative IT-Lösungen

Vattenfall GmbH

ZAB – ZukunftsAgentur Brandenburg

Tagungsleitung

Gesamtleitung:	Petra Hofstedt, BTU
	Ingo Schmitt, BTU
Tutorien und Workshops:	Douglas Cunningham, BTU
	Klaus Meer, BTU
Tag der Informatik:	Jörg Nolte, BTU
Industrie-/Wirtschaftsprogramm:	Rolf Kraemer, IHP Frankfurt Oder und BTU
Studierendenprogramm:	Heinrich Theodor Vierhaus, BTU
Doktorandenprogramm:	Heinrich Theodor Vierhaus, BTU
	Ansgar Scherp, Leibniz-Informationszentrum Wirtschaft und Universität zu Kiel
	Steffen Becker, TU Chemnitz
Sponsoring:	Rolf Kraemer, IHP Frankfurt Oder und BTU
	Ingo Schmitt, BTU
	Petra Hofstedt, BTU
Lokale Organisation:	Claus Lewerentz, BTU
	Steffen Helke, BTU
	Petra Hofstedt, BTU
Öffentlichkeitsarbeit:	Gerd Wagner, BTU
	Douglas W. Cunningham, BTU

GI-Organisationsteam

Cornelia Winter
Elena Kerkmann

Inhalt

Keynote Vorträge

Oliver Deussen
Visual Computing: Von bunten Bildern zu quantifizierbaren Verfahren . . . **27**

Stefan Pappe
Enterprise Hybrid IT **29**

Hasso Plattner
A new architecture for enterprise application software based on in-memory databases **31**

Michael Sonnenschein
Smart Grids – Informatik für die Stromnetze der Zukunft **33**

Energieinformatik

Wilfried Elmenreich
Interoperability Between Smart and Legacy Devices in Energy Management Systems **39**

Holger Hermanns
Energieinformatik im erdnahen Orbit **41**

Hartmut Schmeck
Organic Smart Home – eine flexible Architektur für ebenen- und energieträgerübergreifendes Energiemanagement **43**

Sabine Storandt
Route Planning for Electric Vehicles: Taking Energy Efficiency, Distance, and Reloading Opportunities into Account **45**

IT-Systeme und Architekturen für die Analyse und Gestaltung von Energienetzen

Albrecht Reuter
IKT als Enabler der Energiewende **51**

Jessica Thomsen, Niklas Hartmann, Florian Klumpp, Thomas Erge, Michael Falkenthal, Oliver Kopp, Frank Leymann, Sven Stando, Nino Turek, Christoph Schlenzig, Holger Schwarz
Darstellung des Konzeptes – DMA Decentralised Market Agent – zur Bewältigung zukünftiger Herausforderungen in Verteilnetzen **53**

Oliver Kopp, Michael Falkenthal, Niklas Hartmann, Frank Leymann, Holger Schwarz, Jessica Thomsen
Towards a Cloud-based Platform Architecture for a Decentralized Market Agent **69**

Michael J.M. Wagner, Friedrich Eisenhauer, Dagmar Koß
System Alarming für Energiesysteme **81**

IT und Dienstleistungen für die Energiewende und die Elektromobilität (IDEE)

Carola Gerwig, Dennis Behrens, Helmut Lessing, Ralf Knackstedt
Demand Side Management in Residential Contexts – A Literature Review . **93**

Oliver Norkus, Brian D. Clark, Florian Merkel, Björn Friedrich, Jürgen Sauer, H.-Jürgen Appelrath
An Approach for a Cloud-based Contribution Margin Dashboard in the Field of Electricity Trading **109**

Tobias Weiß, Marco Krause
Auswertung von Energiedaten – Ein Ansatz zur Kombination einer soziologischen Perspektive mit der Methodik des Data Minings **121**

David Bauer, Stefan Kirschbaum, Gregor Wrobel, Julian Agudelo, Philip Voll
Modellbasierte Optimierung von Energiesystemen **137**

Astrid Niee, Michael Sonnenschein
A Fully Distributed Continuous Planning Approach for Decentralized Energy Units **151**

Userinteraktion-Energieoptimierung

Dominic Becking, Jan Budke, Sybille Reichart
Microcurricula in augmentierter Realität! Ein Prototyp zur Hilfe bei der Interaktion mit intelligenten Gebäuden **171**

Andreas Sawatzki, Uwe Sedlak, Oliver Wetter, Uwe Weitkemper, Martin Hoffmann
Der moderne Brandschutz – Fluchtwege der Zukunft **181**

Aljoscha Pörtner, Dennis Sprute, Alexander Weinitschke, Matthias König
Integration of a fall detection system into the intelligent building **191**

Tobias Weiß, Madlen Diesing, Marco Krause
Die Wirkung von Feedback und Goal Setting auf den Energieverbrauch von Privathaushalten – Eine Meta-Analytische Untersuchung **203**

Betriebliche Umweltinformationssysteme (BUIS)

Volker Wohlgemuth
Ein Überblick ber Einsatzbereiche von betrieblichen Umweltinformationssystemen (BUIS) in der Praxis **223**

Johannes Boß, Volker Wohlgemuth
Integration und Weiterentwicklung bestehender Energiemanagement-Applikationen mit dem OpenResKit-Framework **239**

Svetlana Bloching, Jürgen Neumann, Iris Rabener, Ingo Rau
Praxisbeispiel Deutsche Bahn AG: FINK – mit BPM komplexe Fachprozesse im Umweltschutz managen **249**

Richard Rößler, Helen Lachmann, Werner Esswein
Documentational aspects of integrated management systems on the example of a plastics manufacturing company **259**

Heiko Thimm
A Monitoring Framework for Guidance and Risk Control Assistance of Environmental Compliance Officers **271**

Andreas Moeller
Zwecke von betrieblichen Umweltinformationssystemen im Rahmen einer nachhaltigen Entwicklung **281**

Jorge Marx Gómez, Frank Medel-González
Integrating IT-for-Green Project through Web Services with the System of Sustainability Performance Evaluation **291**

Viktor Dmitriyev, Jorge Marx Gómez, Manuel Osmers
Big Data Inspired Water Management Platform for Sensor Data **303**

Zhongkai Li, Jorge Marx Gómez
Modeling for sustainable product development strategies with general morphological analysis **311**

Umweltinformatik zwischen Nachhaltigkeit und Wandel (UINW 2015)

Malte Lärz
Grundkonzeption einer 'Kritischen Theorie der Hybridität' und Implikationen für "nachhaltige Wissenschaften" **325**

Florian Stiel, Frank Teuteberg
On the Environmental Footprint of an IS Conference **335**

Eva Kern
Kommunikationsmöglichkeiten für Aspekte der Nachhaltigkeitsinformatik . **345**

Dominik Trumm, Andreas Filler, Eva Kern, Stefan Naumann
SALT – Situation-sensitive Sustainable Service and Product Alternatives: Vision, Conceptual Application & Challenges **355**

Patrick Piltz, Volker Wohlgemuth
Konzeption und Entwicklung eines Genehmigungsmanagements für KMU unterstützt durch die Augmented Reality **365**

Hans-Knud Arndt
Ist ein Perspektivenwechsel der Umweltinformatik durch ein mögliches Verschwinden der Grenze zwischen physischem Objekt und Software notwendig? **375**

Michael Sonnenschein, Ute Vogel, Christian Hinrichs
From Ecological Modeling to Decentralized Optimization of Smart Power Grids **385**

Pawel Bartoszczuk
Environmental and Energy Informatics: Polish Approaches in Eco-innovation **397**

Digitalisierung und Informationssysteme der Nachhaltigkeitsberichterstattung - Wie Institutionen IKT und Software- Werkzeuge einsetzen, um Kommunikation und Nachhaltigkeitsleistung zu verbessern

Matthias Mokosch, Torsten Urban, Hans-Knud Arndt
A new Visualization Approach for Sustainability **413**

Christina Niemöller, Niclas Bärtling, Oliver Thomas
Nachhaltigkeit durch Hybride Wertschöpfung – Entwicklung eines Reifegradmodells **427**

Volkan Gizli, Jorge Marx Gómez, Simon Stephane Ntomthe Tiako
Nutzung individuell bestimmbarer KPIs im Incident Management zur nachhaltig effektiven Schwachstellen- und Stärkenidentifizierung von IT-Service Landschaften **443**

Bettina Horlach, Paul Drews, Ingrid Schirmer
Enterprise Architecture as a Tool for Managing Corporate Social Responsibility **453**

Ralf Isenmann, Sassen Remmer, Sascha Zinn
Digitalisierung und IKT in der Nachhaltigkeitsberichterstattung von Hochschulen in Deutschland – terra incognita? **469**

Andreas Mladenow, Christine Strauss
Active Sourcing, Matching und Nachhaltigkeitskommunikation beim eRecruiting **483**

Die Automatisierung von Compliance in Cloud-Umgebungen (ACCU)

Andreas Jede, Frank Teuteberg
Evidente Ausfallrisiken im Cloud-Markt – Eine quantitative Analyse der Finanzberichte von Anbietern **499**

Sebastian Luhn, Maxmilian Hils
Ein Rahmenwerk für Datenschutz-Metriken in der Cloud **511**

Bernd Jäger, Reiner Kraft, Annika Selzer, Ulrich Waldmann
Werkzeuge zur Messung der datenschutzkonformen Einhaltung des Verarbeitungsstandorts in der Cloud **525**

Johanna M. Hofmann
Dynamische Zertifizierung von Cloud Computing- Diensten: Eine rechtswissenschaftliche Betrachtung am Beispiel "Verfügbarkeit" **539**

Cyberwarfare

Annegret Bendiek, Tobias Metzger
Deterrence theory in the cyber-century **553**

Ingo Ruhmann
Neue Ansätze für die Rüstungskontrolle bei Cyber-Konflikten **571**

Thomas Reinhold
Möglichkeiten und Grenzen zur Bestimmung von Cyberwaffen **587**

Sven Herpig
Strategic Operations in the Cyber Domain and their Implications for National Cyber Security **597**

SAG WAS – Studentische AusbildunG und berufliche Weiterbildung in AnforderungsSpezifikationen

Ulrike Friedrich, Karin Schlör
Just-in-Time Teaching – Interne Ausbildung 2.0 . . . . . **613**

Veronika Thurner, Axel Böttcher, Kathrin Schlierkamp, Daniela Zehetmeier
Anforderungen an Anforderungen – Vermitteln von Requirements Engineering **617**

Yvonne Sedelmaier, Dieter Landes
Ein mehrstufiges Lehrkonzept für RE **619**

Gero Wedemann
Requirements Engineering an der Hochschule in einer Blockwoche – Best Practices und Lessons Learned **621**

Ralf Reißing
Lehren und Prüfen im Requirements Engineering am Beispiel eines Master-Kurses für Ingenieure **623**

Hochschule2025

Sören Lorenz, Frank Klapper
Projektflut – geht der CIO baden? Multiprojektmanagement an Hochschulen **629**

Thomas Hildmann, Christopher Ritter, Nico Suhl
tubIT Live – eine komplette Studierendenumgebung "to Go" **639**

Markus von der Heyde, Andreas Breiter
Wer entscheidet? IT-Governance an Hochschulen **651**

Christian Erfurth, Ivonne Erfurth
IT-Dienste im Kontext von Hochschulprozessen – Eine IT-Landkarte zur Unterstützung des IT Servicemanagements **663**

Gerald Lach, Mirjana Lach, Erhard Zorn
Integrierte Ressourcen-Planung an Hochschulen **675**

Stefan Wesner, Claudia Pauli, Jens Kohlmeyer, Beate Quester-Brüning, Marcus Link
campus4uulm – Flexibles und integriertes Campus-Management mit Standardsoftware? **687**

André Kleinschmidt
Strategisches zu Campusmanagementsoftware **701**

Andreas Drescher, Michael Meier, Andreas Oberweis, Frederic Toussaint
Systematische Prozessverbesserung mittels präferenzorientierter Ressourcenallokation am Beispiel einer Tutorien-platzvergabe **713**

Patrick Bittner
Modulempfehlungen über Vorgänger- und Nachfolgemodule **725**

Thomas Jankowski, Ulrike Lucke
Effizientes und bedarfsgerechtes Berichtswesen auf Basis eines modularen Systems aus Open-Source-Produkten **735**

Strategisches Informationsmanagement und IT-Governance) im Öffentlichen Sektorâ (SIMÖS)

Michael Schaaf, Franziska Jahn, Kais Tahar, Christian Kücherer, Alfred Winter und Barbara Paech
Entwicklung und Einsatz einer Domänenontologie des Informationsmanagements im Krankenhaus **753**

Konrad Walser, André Meister, Etienne Huber
Tiefenanalyse zum Performance Management in Schweizer Krankenhäusern auf Basis dreier Fallstudien **767**

Christian Kücherer, Manuel Jung, Franziska Jahn, Michael Schaaf, Kais Tahar, Barbara Paech, Alfred Winter
System Analysis of Information Management **783**

Konrad Walser, Philipp Christen, Sabrina Maag, Andrin Regli
Business-IT-Alignment in Gemeinden â Qualitative Forschung anhand dreier größerer Berner Gemeinden **797**

Business Process Management im Öffentlichen im Sektor (BPMÖS)

Silke Holzmüller-Laue und Kati Häger
Evaluation einer Methodik zur standardisierten Abbildung von Verwaltungsprozessen **819**

Simon Enkerli, Konrad Walser
Entwicklung eines generischen Gemeindereferenzprozessmodells in Anlehnung an das ARIS-Modell **835**

Digital Enterprise Architectures

Kurt Sandkuhl
Impact of Cyber-Physical System Implementation on Enterprise Architectures: A Case Study **855**

Dierk Jugel, Stefan Kehrer, Christian M. Schweda, Alfred Zimmermann
A Decision-Making Case for Collaborative Enterprise Architecture Engineering **865**

Michael Möhring, Rainer Schmidt
Daten-getriebene Unternehmensarchitekturen im E-Commerce für das präventive Retourenmanagement **881**

Felix Timm, Matthias Wißotzki, Christina Köpp, Kurt Sandkuhl
Current State of Enterprise Architecture Management in SME Utilities . . . **895**

Manuel Breu, Klaus Berndl, Thomas Heimann, Lars Santesson
*i*Gov a Feeling – Herausforderungen und Technologien zur Realisierung eines Studienportal für das interaktive Heute* **909**

Daniela Fisseler, René Reiners
Prozessorientierte überwachung in der Produktion **917**

Workshop zum Stand, den Herausforderungen und Impulsen des Geschäftsprozessmanagements

Christoph Sigmanek, Birger Lantow
Modellierung wissensintensiver Geschäftsprozesse aus der Sicht des Wissensmanagements **933**

Anne Gutschmidt, Kurt Sandkuhl
Multi-Touch-Tisch oder Plastikfolie? – Design einer Studie zum Vergleich von Medien in der Modellierung **947**

Michael Gebhart, Marco Mevius, Peter Wiedmann
Wie erleben Anwender ihre Geschäftsprozesse? User Feedback mittels Mobile App **963**

Hasan Koç, Kurt Sandkuhl
From Business Process Models to Capability Models **979**

Oliver Arnold, Ralf Laue
Überblick und Weiterentwicklung von Methoden zur Beherrschbarkeit von Variabilität in Geschäftsprozessmodellen **993**

Workshop zur Gesundheitstelematikinfrastructure so wie zu ethischen, juristischen und sozialen Aspekten informatischer Systeme in Pflege und Gesundheitsversorgung (GTU ELSA)

Marc Walterbusch, Tim Hoffmann und Frank Teuteberg
Haltung und übertragung von Patientendaten im Cloud Computing: Anforderungserhebung und prototypische Implementierung eines Verschlüsselungsframeworks **1011**

Cordula Endter
How Older People Matter – Nutzer_innenbeteiligung in AAL-Projekten . . . **1027**

Deklarativen Modellierung und effizienten Optimierung komplexer Probleme (MOC)

Peter Sauer, Denny Schneeweiß, Petra Hofstedt
Divide & Conquer – Komplexe Constraint-Probleme durch Aufteilen lösen . **1047**

Michaela Helbig, Jens Weller, Hajo Wiemer und Gunnar Dietz
Über den Umgang mit modellabhängigen Komponenten: Verfeinerungen und ihre Auswirkungen **1061**

Anna Prenzel
Ein Modell zur Beschreibung des Interaktionsablaufs zwischen Mensch und Computer bei der Lösung von Planungsproblemen **1077**

Hans Schlenker
Produktionsplanung mit einer kombinierten Lösung aus mathematischer Optimierung und anderen Business Analytics Werkzeugen **1093**

Hans-Joachim Goltz und Armin Wolf
Reservierungsmanager für Cloud-Ressourcen mittels constraintbasierter Programmierung **1097**

Ulrich John
Digitales Unternehmen – Bausteine für Effizienz, Agilität und Transparenz . **1103**

Thomas Kuhnke und Tom Hausten
Constraint-basierte Planung und Optimierung von Prüfungsterminen mithilfe einer graphischen Benutzeroberfläche **1119**

Dawn Domaschk and Florin Popescu
Constrained Production Schedule Optimization of Output- Normalized Expenditures under Uncertainty in Shift Duration and Energy Price Forecasts **1125**

Sven Löffler
Eine grafische, interaktive Benutzeroberfläche für den Constraint-basierten Produktkonfigurator FdConfig **1131**

Petra Hofstedt, Walter Hower, Ulrich John, Andreas Kmoch, Armin Wolf
Deklarative Modellierung und effiziente Optimierung – quo vadis? Ausgewählte Herausforderungen und Potenziale für die nächsten Jahre – zur Diskussion **1135**

IT-Unterstützung in Emergency Management & Response

Thomas Pappert, Florian Brauner, Ompe Aimé Mudimu, Alex Lechleuthner, Konrad
IT-gestütztes Informationsmanagement in grenzüberschreitenden Großschadensereignissen **1143**

Frank Geyer, Dr. Volkmar Schau, Prof. Dr. Wilhelm R. Rossak
Social Media im Krisen- und Katastrophenmanagement – Ein Überblick . . **1155**

Big Data, Smart Data and Semantic Technologies (BDSDST)

Nicolai Krüger, Frank Teuteberg
From Smart Meters to Smart Products: Reviewing Big Data driven Product Innovation in the European Electricity Retail Market **1171**

Raik Niemann and Todor Ivanov
Modelling the Performance, Energy Consumption and Efficiency of Data Management Systems **1183**

Samaneh Beheshti-Kashi, Michael Lütjen, Lennard Stoever, Klaus-Dieter Thoben
TrendFashion – A Framework for the Identification of Fashion Trends . . . **1195**

Artem Zakharchuk, Sergey Kovalchuk, Denis Nasonov
Dynamic Domain-Specific Language for BigData Tasksâ Description **1207**

Thomas Göpfert, Andreas Breiter
Knowledge Discovery in Big Data: Herausforderungen durch Big Data im Prozess der Wissensgewinnung am Beispiel des CRISP-DM **1219**

Wolfgang Orthuber
How to make quantitative data on the web searchable and interoperable part of the common vocabulary **1231**

Cognitive and Dependable Systems

Karsten Weber, Alena Wackerbarth
Partizipative Gestaltung verlässlicher Assistenzsysteme für die Pflege alter und hochbetagter Menschen **1247**

Ute Schmid
Cognitive Systems: Goals, Approaches, Applications **1251**

Michael Breuß, Douglas W. Cunningham and Martin Welk
Scale Spaces for Cognitive Systems: A Position Paper **1253**

Günther Wirsching
Mathematical structures for modeling semantics, uncertainty, and cognitive processes **1257**

Douglas W. Cunningham
Perceptual Graphics: The symbiosis between and formal similarity of perceptual psychology and computer graphics **1259**

David Zellhöfer
Einsatzmöglichkeiten kognitiver Retrieval-Modelle in Digitalen Bibliotheken **1261**

Gamification und Virtuelle Welten

Andreas Mladenow, Niina Maarit Novak, Christine Strauss
Micropayments in virtuellen Welten â Prozessmodell und Nutzung bei 7- bis 12-jährigen Besuchern **1267**

Christian Karl Grund
How Games and Game Elements Facilitate Learning and Motivation: A Literature Review **1279**

Linda Eckardt, Jan-Paul Huttner, Susanne Robra-Bissantz
GamEducation in einer virtuellen 3D-Umgebung mit Googles Virtual-Reality-Brille Cardboard **1295**

Teilhabe an der allgegenwärtigen Kommunikation

Susanne Dirks, Nadezda Anton, Christian Bühler
Förderung der medialen Teilhabe von Menschen mit erworbenen Hirnsch Ìädigungen durch assistive Technologien **1311**

Wiebke Köhlmann, Nils Dressel, Dustin Wegner
Erweiterung eines virtuellen Klassenzimmers zur Verbesserung der Zugänglichkeit für Blinde **1325**

Christian Bühler, Christian Radek und Arbeitsgruppe der TU Dortmund
Meldestelle fr digitale Barrieren **1341**

Ricardo Buettner
Inklusion in Social Media: Die Perspektive der Betroffenen **1349**

Mathias Haimerl
Development towards a generally applicable process to inspect and verify accessibility of web pages **1361**

Fehlertolerante und energieeffiziente eingebettete Systeme: Methoden und Anwendungen

Goran Panić
Low Power Design Techniques **1379**

Milos Krstic
Reducing Power Consumption in Fault Tolerant ASICs **1381**

Roland Jancke, Christoph Sohrmann, Kay-Uwe Giering
Fehler vermeiden oder tolerieren? Methoden zur Berücksichtigung der Zuverlässigkeit im Entwurf integrierter Schaltungen **1383**

Kai-Uwe Irrgang, Thomas B. Preußer, Rainer G. Spallek
Kompression von Tracedaten auf Bitebene basierend auf einem LZ77-Wörterbuchansatz **1385**

Christian Gleichner, Heinrich T. Vierhaus
Test eingebetteter Prozessoren im Zielsystem mit hoher diagnostischer Auflösung **1399**

Gökçe Aydos, Görschwin Fey
Parity-based Soft Error Detection with Software-based Retry vs. Triplication-based Soft Error Correction – An Analytical Comparison on a Flash-based FPGA Architecture **1415**

Davide Dicorato, Heinrich T. Vierhaus
Detection and Correction of Logic Errors Using Extra Time Slots **1431**

Günther Nieß, Thomas Kern, Michael Gössel
Möglichkeiten der Modellierung von Fehlern in MLC-Flash-Speichern durch Fehlergraphen **1445**

Patryk Skoncej, Felix M ühlbauer, Mario Schölzel
Softwarebasierte Fehlertolerenz für Flash-Speicher von mikrocontroller-basierten Systemen **1461**

Frank Vater, Goran Panic, Mario Schölzel
Möglichkeiten der Nutzung von RRAM in Low-Power Microcontrollern . . . **1475**

Industry 4.0: Informatics enables new Production Systems

Peter Liggesmeyer
Big Data, Industrie 4.0 und Co.: Wie sich die Welt verändert **1491**

Claus Hilger
Herausforderung:Vertikale Integration – Vom Shop Floor zum Top Floor . . **1493**

Jürgen Jasperneite
Smart Factory based on intelligent technical systems **1495**

Frank Ellinger
Fast Actuators Sensors and Transceivers (FAST) **1497**

Hans D. Schotten
5G und Industrie 4.0 . . . . . **1499**

Ulrich Berger, Kornelius Wächter und Janny Klabuhn
Einfluss von Industrie 4.0 auf Wertschöpfungsnetze am Beispiel der Textilreinigungsbranche **1501**

Jörg F. Wollert
Generische Funksysteme für IoT und Industrie 4.0 . . . . . **1509**

Armin Wulf, Wolfgang Endemann und Rüdiger Kays
Entwurf und Auslegung eines Funksystems für dichte Netzwerke in der Fabrikautomatisierung **1521**

Johannes Ax, Aurel Buda, Daniel Schneider, John Hartfiel, Lars Dürkop, Thorsten Jungeblut, Jürgen Jasperneite, Andreas Vedral und Ulrich Rückert
Universelle Echtzeit-Ethernet Architektur zur Integration in rekonfigurierbare Automatisierungssysteme **1533**

Johannes Reich
Eine semantische Klassifikation von Systeminteraktionen **1545**

Maximilian Speicher, Kristina Tenhaft, Simon Heinen and Harry Handorf
Enabling Industry 4.0 with holobuilder **1561**

Automotive Software Engineering

Markus Bechter
Softwareplattform für zukünftige automotive Anwendungen **1581**

Martin Sükraut, André Schmitt, Jörg Kaienburg
Safety Functions on Commodity Hardware with Diversified Encoding . . . **1583**

Bernhard Kaiser, Stefan Sonski, Suryo Buono, Hauke Petersen, Justyna Zander
Lightweight Contracts for Safety-Critical Automotive Systems **1595**

Thomas Noack, Steffen Helke
Automatische Verlinkung von Anforderungen und Testfällen–Fallbasierte Filterung **1609**

Hans-Werner Wiesbrock
Eine Stochastische Testmethodik für den Robustheitstest Automobiler Steuergeräte **1621**

Patrick Helmholz, Susanne Robra-Bissantz
AmbiTune: Kontextorientierte Musikeinspielung im Fahrzeug **1635**

Jean-Pierre Seifert, Michael Peter
Automotive MILS **1647**

Christian Dernehl, Norman Hansen, Thomas Gerlitz, Stefan Kowalewski
Static Value Range Analysis for Matlab/Simulink-Models **1649**

Quang Minh Tran, Jonas Winkler, Christian Dziobek
SLRefactor: Ein Refactoring-Ansatz fr Simulink-Modelle **1661**

Tobias Wägemann, Albert Werner
Generating Multi-objective Programs from Variant-rich EAST-ADL Product Line Architectures **1673**

Bernhard Rumpe, Christoph Schulze, Johannes Richenhagen, Axel Schloßer
Agile Synchronization between a Software Product Line and its Products . . **1687**

Peter Schiele, Alexander Siller
Qualitätssicherung in der Software-Entwicklung am Beispiel des BMW i3 . **1699**

Tutorials

Stefan Schleyer
Remanufacturing in the Information Age **1713**

Detlef E. Schröder
Angewandte Statistik zur Analyse von Nutzerdaten in der Energiewirtschaft **1715**

Leif Oppermann, Elke Hinrichs, Ulrich Schade, Thomas Koch, Manuela Rettweiler, Frederike Ohrem, Patrick Plötz, Carsten Beier, Wolfgang Prinz
EnArgus: Zentrales Informationssystem Energieforschungsförderung **1717**

Katja Stecklina, Clemens Passeck
Grundlagen sicherheitsgerichteter Software-Entwicklung **1731**

Doktorandenprogramm

Malte Zuch
Effizienz im elektromobilen Massenmarkt **1739**

Tim Waage
Durchsuchbare Verschl usselung in NoSQL Datenbanken **1747**

Johannes Hiltscher
Hardware supported efficient accelerator partitioning for workstation consolidation and virtualization **1759**

Timo Hönig
From Sensibility to Responsibility: The Impact of System Software on Ecological and Economical Sustainability of Computing Systems **1771**

Khalil Omar
Towards Improving the Usability of Mobile ERP. A Model for Devising Adaptive Mobile UIs to Improve the Usability of Mobile ERP **1783**

Arthur Vetter
Vermeidung von IT Störungen unter Einsatz eines Kontrollsystems zur Durchführung von IT Changes **1795**

Studierendenkonferenz Informatik(SKILL)

Susanne Kenner, Raphael Thaler
Eine Smart Grid Architektur und Umsetzung zwecks Datenanalyse und Leistungsmanagement am Beispiel des Stromnetzes der OTH Regensburg **1805**

Sebastian Wolff
Building A State-Of-The-Art Model Checker **1817**

Tho Le
A recommended framework for anomaly intrusion detection system (IDS) . **1829**

Andreas Stadelmeier
Java type inference as an Eclipse plugin **1841**

Lea Kristin Gerling
Evaluation der Leistungsfähigkeit ausgewählter Mutationstestwerkzeuge . . **1853**

Tim Mühlsteff, Andreas Schäfer
Patterns im Umfeld prozessgetriebener Anwendungen zur Dokumentenverarbeitung mit BPMN **1865**

Lars Schütz
Game Event Lenses: Focus+Context Visualizations of Computer Game Data **1877**

Jan Asmuth, Markus Bleß, Doreen Brunner, Mark Deppe, Marcel Prügel
Entwicklung eines Spieles zum Thema Schatten-IT **1889**

Tobias Mueller
Framework for Fuzzing USB stacks with Virtual Machines **1901**

Sebastian Runge
Mechanismus zur Synchronisation verteilter Datenmodelle für kollaborative Editoren **1913**

Lukas Eipert
Metadatenextraktion und Vorschlagssysteme im Visual SPARQL Builder . . **1925**

Paul Jähne
Erzeugung minimaler Spannbäume auf ungerichteten, kantengewichteten Graphen mit den Algorithmen von Kruskal, Prim und Boruvka **1937**

Ronald Smith Djomkam Yotedje
GIET: Generic Information Extraction using Triple Store Databases **1948**

Tag der Informatik

30. September 2015

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Visual Computing: Von bunten Bildern zu quantifizierbaren Verfahren

Oliver Deussen
Universität Konstanz

Abstract: Visual Computing fasst Methoden der Bildanalyse, Computergrafik, Visualisierung sowie Mensch-Maschine Interaktion zusammen mit dem Ziel, komplexe Daten aus Gesellschaft, Wirtschaft und Wissenschaft dem Menschen visuell zugänglich zu machen. An mehreren konkreten Beispielen wird dies veranschaulicht, es werden Probleme der Bildmanipulation aufgezeigt und die notwendige Quantifizierung von Visualisierungsverfahren besprochen.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Enterprise Hybrid IT

Stefan Pappe
IBM

Abstract: New Cloud, Analytics, Mobile, and Social solutions drive digital disruption across all industries, and accelerate business cycles. To deliver innovative products and services quickly in such an environment, enterprises need to become composable businesses, able to decompose existing processes and recombine them with new ones. IT services supporting such a business design need to be both flexible and robust, and they need to span different delivery environments, from enterprise data centers to clouds. They need to enable agile deployment of new services and deriving value from existing ones. They need to be open and expandable to embrace eco-systems and value chains. They need to combine client facing Systems of Engagement with enterprise Systems of Records into Systems of Insight that allow the delivery of optimized services that are both personalized and globally scalable. We call this approach to IT service delivery Enterprise Hybrid IT.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

A new architecture for enterprise application software based on in-memory databases

Hasso Plattner
SAP AG

Abstract: Six years ago, I proposed a common database architecture for transaction processing and analytical systems using a columnar in-memory database, disputing the common belief that columnar storages are not suitable for transactional workloads in enterprise applications.

Today, the concept has been widely adopted in academia and industry and it is proven that it is not only feasible but also lead to a dramatic simplification of data models, a reduction in transaction complexity, the highest flexibility of queries, and finally an impressive performance at lower Total Cost of Ownership. This redefines the way we build enterprise software because we are now able to run analytical queries on large data sets directly on the finest granularity of data, eliminating the need to maintain pre–built aggregate tables. So, the next generation of applications is possible and realizes, for example, interactive boardroom meetings or personalized medicine.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Smart Grids – Informatik für die Stromnetze der Zukunft

Michael Sonnenschein
Carl von Ossietzky Universität Oldenburg

Abstract:

Eine zentrale Herausforderung der Energiewende besteht in der Integration einer Vielzahl dezentraler Stromerzeuger, steuerbarer Verbraucher und Speicher zu einem langfristig funktionsfähigen, robusten Versorgungssystem. Hierzu sind Methoden der Informatik unverzichtbar: Auf der einen Seite müssen heterogene Teilsysteme über Standards zu einem interoperablen Gesamtsystem integriert werden. Hierzu wurden bereits standardisierte Architekturmodelle wie die SGAM entwickelt. Auf der anderen Seite ist eine algorithmische Unterstützung von Koordinationsaufgaben erforderlich. Hierzu zählen der jederzeitige Abgleich von Angebot und Nachfrage, die fahrplanbasierte Stromlieferung durch eine große Zahl kleiner Anlagen, wie auch die Stabilisierung von Spannung und Frequenz. In künftigen Versorgungssystemen mit einer Vielzahl individueller Betreiber von Anlagen ist eine naheliegende Lösung die Entwicklung verteilter Verfahren für diese Zwecke.

Nach einer Einführung in die Thematik werden in diesem Vortrag zunächst verteilte Verfahren zur Steuerung virtueller Kraftwerke als Beispiel für den Einsatz von Informatik-Methoden in der künftigen Stromversorgung vorgestellt. Darauf aufbauend wollen wir aber auch Fragen der Netzstabilität und die daraus resultierenden Herausforderungen und Lösungsansätze aus der Informatik diskutieren.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Workshop Energieinformatik

Workshop Energieinformatik

Martin Leucker[1], Martin Sachenbacher[2] und René Schönfelder[3]

Abstract: Im Rahmen dieses halbtägigen Workshops werden aktuelle Themen des Gebiets der Energieinformatik von führenden Forscherinnen und Forschern aus dem deutschsprachigen Raum vorgestellt. Der Themenschwerpunkt liegt dabei auf der formalen Behandlung von Aspekten der Energieinformatik, insbesondere hinsichtlich mathematischer Modellierung, formaler Analyse und Algorithmik.

Die Energieinformatik ist ein wachsender Bereich der Informatik, in welchem Aspekte von Informatik und Energie behandelt werden, mit einem gewissen Fokus auf Elektromobilität, Energiespeicherung und Smart Grids. Dabei geht es einerseits um technische Aspekte wie nachhaltige Energieerzeugung durch Windkraft oder Solartechnik und Stromspeicherung in Elektrofahrzeugbatterien; andererseits werden ökonomische Aspekte wie die Fragestellung untersucht, wann es sich für einen Fahrer lohnt, Strom ins Netz zurückzuspeisen.

Um diese Ziele zu erreichen, baut die Energieinformatik auf einer Reihe grundlegender Theorien, Methoden und Werkzeuge der Informatik und verwandter Gebiete auf:

- Modellierungstechniken für komplexe, hybride dynamische Systeme: maschinelles Lernen aus Daten, stochastische Modelle, Constraintnetze, temporale Logik, hybride Automaten, etc.
- Modellgestützte Simulation, Verhaltensvorhersage und Analyse: numerische und qualitative Simulation, modellbasiertes Schließen, Constraint Reasoning, Zustandsschätzung, Verifikation, Diagnose, Monte-Carlo- und Sampling-basierte Verfahren, etc.
- Berechnung ökonomischer Gleichgewichte und Strategien für interagierende Agenten: Multiagenten-Systeme, Adversarial Reasoning, Spieltheorie, kombinatorische Auktionen, elektronische Märkte, etc.
- Techniken für die Optimierung, Planung und Steuerung in technischen Systemen: Lineare Programmierung, Constraint Programming, Autonomous Planning and Scheduling, Supervisory Control, Fault-tolerant Control, etc.
- Technische Umsetzung und Implementierung der Verfahren: Controller, eingebettete Systeme, Cyber-physische Systeme, etc.

Diese Methoden und Techniken können in verschiedenen Anwendungsgebieten der Energieinformatik eingesetzt werden, insbesondere:

- Mobilität: IT-gestützte Elektromobilität, vernetzte Fahrerassistenzsysteme, energieeffizientes Routing, Verkehrsmanagement- und Logistiksysteme, etc.

[1] Universität zu Lübeck, Institut für Softwaretechnik und Programmiersprachen, Ratzeburger Allee 160, 23562 Lübeck, leucker@isp.uni-luebeck.de
[2] LION Smart GmbH, Dieselstraße 22, 85748 Garching, martin.sachenbacher@lionsmart.com
[3] Universität zu Lübeck, Institut für Softwaretechnik und Programmiersprachen, Ratzeburger Allee 160, 23562 Lübeck, schoenfr@isp.uni-luebeck.de

- Energienetze: Integration von Wind- und Solarkraft in Stromnetzen, virtuelle Kraftwerke, Demand Side Management/Response, Vehicle-to-Grid, etc.
- Smart Homes/Buildings: intelligente Gebäudesteuerung, Smart Metering, Autonome Mikro-Grids, etc.
- Grüne IT: intelligente Rechen- und Datenzentren, Energiemanagement in der IuK-Technologie, etc.

Das Programm des halbtägigen Workshops besteht aus folgenden eingeladenen Vorträgen (in alphabetischer Reihenfolge), die jeweils über ein Gebiet einen Überblick bieten.

- **Wilfried Elmenreich** – Interoperability Between Smart and Legacy Devices in Energy Management Systems
- **Holger Hermanns** – Energieinformatik im erdnahen Orbit
- **Hartmut Schmeck** – Organic Smart Home – eine flexible Architektur für ebenen- und energieträgerübergreifendes Energiemanagement
- **Sabine Storandt** – Route Planning for Electric Vehicles: Taking Energy Efficiency, Distance, and Reloading Opportunities into Account

Organisation: Martin Leucker (Universität zu Lübeck), Martin Sachenbacher (LION Smart GmbH, Garching), René Schönfelder (Universität zu Lübeck)

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Interoperability Between Smart and Legacy Devices in Energy Management Systems

Wilfried Elmenreich[1]

Abstract: Energy management systems can help to decrease energy consumption by giving user feedback or by directly controlling devices. Smart appliances create a network of devices that can be addressed and controlled via a defined network interface. However, legacy devices will establish a significant portion of a system's power consumption and, therefore, need to be included into the management system. We propose an open architecture to integrate smart and non-smart devices by using smart plugs and non-intrusive load monitoring methods. Devices are connected either as (i) smart appliances via a fieldbus or wireless network, (ii) legacy devices connected to a smart plug, or (iii) other legacy devices being detected from a time sequence of power consumption values, which are disaggregated into the power draws of different devices. At a service layer, device properties are presented in a unified way including a machine-readable description of their features and properties. The data layer provides an abstract representation of data and functionalities. It connects to the application layer where different applications can access the data. The system supports mechanism for service discovery, service coordination, and service and resource description.

[1] Universität Klagenfurt, Institut für Vernetzte und Eingebettete Systeme, Universitätsstraße 65-67, 9020 Klagenfurt am Wörthersee, Österreich, wilfried.elmenreich@aau.at

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Energieinformatik im erdnahen Orbit

Holger Hermanns[1]

Abstract: Dieser Vortrag beleuchtet Aspekte der Energieinformatik anhand einer Serie von Nano-Satelliten, die aktuell im All oder auf dem Weg dorthin sind. Hersteller und Betreiber ist das dänische Unternehmen GomSpace. Auf 2 oder 3 Liter Volumen tragen diese Satelliten verschiedene elektrische Apparaturen, zum Beispiel zum Lokalisieren und Verfolgen von kommerziellen Flugzeugen. Der Energiebedarf wird dabei durch eine Anzahl Solarmodule und eine Batterie von 5 Amperestunden Kapazität gedeckt. Allerdings ist es per se nicht offenkundig, welche Aufgaben wann erbracht werden können, ohne die Batterie über Gebühr zu strapazieren oder eine kritische Tiefentladung zu riskieren.

Der Vortrag führt in die Problematik ein, und beleuchtet, welche Arten von Modellen bei der kurz- und mittelfristigen Missionsplanung zum Einsatz kommen, insbesondere in Bezug auf zukünftige „Satellite-as-a-Service"-Missionen, bei denen zum Zeitpunkt des Launches nur die prinzipiellen Möglichkeiten des Satelliten festliegen, und als Dienstleistungen vermarktet werden sollen.

Diese Forschung wird gemeinsam mit Dr.J̃an Krcal und Gilles Nies durchgeführt. Sie wird von der EU im Rahmen des FP7-Projektes SENSATION (318490) gefördert.

[1] Saarland University, Department of Computer Science, Campus Saarbrücken, 66123 Saarbrücken, hermanns@cs.uni-saarland.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Organic Smart Home – eine flexible Architektur für ebenen- und energieträgerübergreifendes Energiemanagement

Hartmut Schmeck[1]

Abstract: Mit dem Organic Smart Home (OSH) ist im Verlauf mehrerer Forschungsprojekte am KIT eine flexible Architektur entstanden, die ein effektives und selbstoptimierendes Energiemanagement in verschiedenen Szenarien ermöglicht. Auf der untersten Ebene erlaubt ein modularer Ansatz eine optimierende Einsatzplanung energierelevanter Komponenten in einem Haushalt, einem Gebäude oder einem anderen Objekt, bei dem die Lastprofile und Freiheitsgrade der Komponenten bestmöglich ausgenutzt werden, um einen Ablaufplan gemäß den aktuellen Anforderung des Energienetzes zur erzeugen. Das OSH kann sowohl zum Energiemanagement in realen Umgebungen eingesetzt werden, als auch zur Simulation des Verhaltens in potentiellen Szenarien. Ein Entity Abstraction Layer in Verbindung mit einem Data Custodian Service erlaubt die angepasste Weitergabe relevanter Informationen an die nächsthöhere Ebene, deren Management wiederum durch ein OSH übernommen werden kann. Der Vortrag erläutert das Konzept des OSH und des damit eng verbundenen Energy Management Panel als Schnittstelle für den jeweiligen Nutzer und präsentiert einige der damit erzielten Forschungsergebnisse, insbesondere auch für ein energieträgerübergreifendes Energiemanagement, das prototypisch bereits im Energy Smart Home Lab des KIT und im FZI House of Living Labs umgesetzt wird.

[1] Karlsruher Institut für Technologie (KIT), Institut AIFB, FZI Forschungszentrum Informatik, KIT-Campus Süd, 76128 Karlsruhe, hartmut.schmeck@kit.edu

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Route Planning for Electric Vehicles: Taking Energy Efficiency, Distance, and Reloading Opportunities into Account

Sabine Storandt[1]

Abstract: Electric Vehicles (EVs) exhibit important advantages compared to fuel driven cars: The necessary energy can be produced from regenerative sources like wind, hydropower, or solar energy. In use, there is the possibility to recuperate energy during deceleration phases or when going downhill. Furthermore EVs typically exhibit lower emissions to their immediate environment in terms of combustion gases or noise levels. On the other hand, current EV technology still suffers from inconveniences which prevent an even faster acceptance of E-mobility: Due to weight and space constraints, EVs only have a limited energy reservoir constraining their cruising ranges. Moreover, reloading typically takes quite a long time, and due to the non-ubiquity of loading stations is not always possible. All these aspects have to be considered when computing reasonable routes for EVs. This poses interesting research tasks as computing energy-optimal paths, or finding shortest paths on which the EV does not run out of energy or reloads at most once. The talk will focus on recent trends and improvements in the area of EV-route planning. This includes methods for placing loading stations in road networks such that EVs can drive on conventional shortest path without range anxiety.

[1] Universität Freiburg, Institut für Informatik, Georges-Köhler-Allee 51, 79110 Freiburg, storandt@informatik.uni-freiburg.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

IT-Systeme und Architekturen für die Analyse und Gestaltung von Energienetzen

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

IT-Systeme und Architekturen für die Analyse und Gestaltung von Energienetzen

Niklas Hartmann[1], Holger Schwarz[2]

Abstract: Der wachsende Anteil erneuerbarer Energien und die fortschreitende Liberalisierung des Strommarktes erfordern einen zukunftsweisenden Ausbau bestehender Energiesystemtopologien. So müssen zukünftige Energiesysteme dynamisch und automatisiert auf Schwankungen im Energieangebot und der Energienachfrage reagieren können. Dafür bedarf es einer Weiterentwicklung von Energienetzen zu sogenannten „Smart Grids“, welche sich dadurch auszeichnen, dass neben der Energieverteilung auch neue Funktionalitäten wie das dynamische Anbinden von Energieerzeugern oder das Abschalten von Verbrauchern nach definierten Optimierungsparametern beherrscht werden. Dafür müssen Daten über Erzeuger, Verbraucher, aber auch andere Flexibilitäten wie Energiespeicher kontinuierlich erfasst und in IT-Systemen verarbeitet werden. Solche Smart-Data-Lösungen bieten auf Basis von sensorbasierten, vernetzten, intelligenten IT-Systemen viele Chancen, ungenutzte Potenziale für den zukünftigen Betrieb der Energiesysteme zu erschließen.
Dieser mit Unterstützung des GI-Fachbereichs Datenbanken und Informationssysteme organsierte Workshop adressiert die Analyse und Gestaltung von Energienetzen mit einem besonderen Schwerpunkt auf den zugrundeliegenden IT-Systemen und deren Architekturen. Insbesondere wird diskutiert, wie die großen Datenmengen, welche in Smart Grids anfallen, zeitnah und zielgerichtet verarbeitet und analysiert werden können. Neben methodischen und algorithmischen Ansätzen wird insbesondere erörtert, wie diese IT-Systeme zukunftsorientiert unter Berücksichtigung neuer Paradigmen und Konzepte aus den Bereichen Cloud Computing, Big Data und Internet of Things gestaltet werden können. Im Kontext dieser IT-Systeme werden auch weitere Aspekte des Smart Grids, wie Optimierungsmodelle für die Planung und Steuerung von Energiesystemen fokussiert.
Der Workshop umfasst zwei Sessions sowohl mit eingeladenen als auch mit eingereichten und begutachteten Beiträgen. Die eingeladenen Beiträge adressieren die Rolle der IKT im Rahmen der Energiewende sowie die Anforderungen die insbesondere aus Sicht der Systembetreiber an die IKT gestellt werden. Damit stehen zunächst die Anforderungen aus der Praxis im Vordergrund. In den weiteren Vorträgen werden dann einzelne Aspekte und Lösungsansätze vertieft. Hierzu gehört die Konzeption einer neuen Marktrolle, eine Cloud-basierten Plattform zur Unterstützung dieser Marktrolle sowie System Alarming in Energienetzen.

Organisation: Niklas Hartmann (Fraunhofer-Institut für Solare Energiesysteme ISE), Holger Schwarz (Universität Stuttgart)

[1] Fraunhofer-Institut für Solare Energiesysteme ISE, Energy Systems and Markets, Heidenhofstraße 2, 79110 Freiburg, niklas.hartmann@ise.fraunhofer.de
[2] Institut für Parallele und Verteilte Systeme, Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, schwarz@ipvs.uni-stuttgart.de

Programmkomitee: Thomas Erge (Fraunhofer ISE), Johann-Christoph Freytag (Humboldt-Universität zu Berlin), Niklas Hartmann (Fraunhofer ISE), Kai Hufendiek (Universität Stuttgart), Wolfgang Lehner (TU Dresden), Frank Leymann (Universität Stuttgart), Bernhard Mitschang (Universität Stuttgart), Dominik Möst (TU Dresden), Albrecht Reuter (Fichtner Consulting AG), Krzysztof Rudion (Universität Stuttgart), Christoph Schlenzig (Seven2one GmbH), Holger Schwarz (Universität Stuttgart), Christoph Weber (Universität Duisburg-Essen)

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

IKT als Enabler der Energiewende

Albrecht Reuter[1]

Abstract: Die Energiewende hat seit dem Beschluss des Deutschen Bundestages über die Energiewendegesetze am 30. Juni 2011 Fakten geschaffen. So steigt der Zubau an erneuerbaren Energietechnologien ungebrochen weiter an, zahlreiche neue Player betreten die Wettbewerbsarena, neue Geschäftsmodelle werden erprobt, die Kostendegression für die neuen Komponenten des Energiesystems entwickelt sich weiter nach unten und die Zahl derer, die aktiv partizipieren möchten steigt. All dies wird nur durch moderne IKT möglich, ohne IKT gibt es keine Energiewende und auch der jüngste G7 Beschluss nach vollständiger Decarbonisierung in diesem Jahrhundert würde nicht umgesetzt werden können.

Keywords: Energiesystem, Energiewende, IKT

[1] Fichtner IT Consulting AG, Sarweystraße 3, 70191Stuttgart, Albrecht.Reuter@fit.fichtner.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Darstellung des Konzeptes – DMA Decentralised Market Agent – zur Bewältigung zukünftiger Herausforderungen in Verteilnetzen

Jessica Thomsen[1], Niklas Hartmann[2], Florian Klumpp[3], Thomas Erge[4], Michael Falkenthal[5], Oliver Kopp[6], Frank Leymann[7], Sven Stando[8], Nino Turek[9], Christoph Schlenzig[10], Holger Schwarz[11]

Abstract: In der vorliegenden Veröffentlichung wird ein Konzept für einen neuen Marktakteur im Strommarkt vorgestellt, der im zukünftigen Smart Grid als Aggregator und Planer fungieren kann. Dieser Decentralised Market Agent – DMA – soll die Informationen aller vorhandenen Erzeugungs- und Speicheranlagen, Lasten und Netzinformationen auf Verteilnetzebene aggregieren sowie mit lokalen Akteuren und an den zentralen Märkten agieren um einen kostenoptimalen Betrieb und Ausbau des Systems Verteilnetzes zu realisieren. Zur Handlungsfähigkeit dieser neuen Marktrolle bedarf es hochauflösender Messungen im Verteilnetz und einer „real-time" Aufbereitung der Messdaten. Im vorliegenden Paper sollen das Konzept sowie die notwendigen Bausteine zur Erreichung der Handlungsfähigkeit des DMA vorgestellt sowie die zukünftig geplanten Untersuchungen erläutert werden. Die detaillierte Entwicklung des Konzepts sowie weiterführende Analysen sind Teil des Projektes NEMAR – Netzbewirtschaftung als neue Marktrolle, gefördert durch BMWi im Rahmen der Forschungsinitiative Zukunftsfähige Stromnetze.

Keywords: DMA, IKT, Optimierung, Flexibilität, Verteilnetz, Datenmanagement

[1] Fraunhofer-Institut für Solare Energiesysteme ISE, Energy Systems and Markets, Heidenhofstraße 2, 79110 Freiburg, jessica.thomsen@ise.fraunhofer.de
[2] Fraunhofer-Institut für Solare Energiesysteme ISE, Energy Systems and Markets, Heidenhofstraße 2, 79110 Freiburg, niklas.hartmann@ise.fraunhofer.de
[3] Fichtner GmbH & Co. KG, Abteilung für Energiewirtschaft, Sarweystrasse 3, 70191 Stuttgart, Florian.Klumpp@fichtner.de
[4] Fraunhofer-Institut für Solare Energiesysteme ISE, Intelligente Energiesysteme, Heidenhofstraße 2, 79110 Freiburg, thomas.erge@ise.fraunhofer.de
[5] Institut für Architektur von Anwendungssystemen, Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, falkenthal@iaas.uni-stuttgart.de
[6] Institut für Parallele und Verteilte Systeme, Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, kopp@ipvs.uni-stuttgart.de
[7] Institut für Architektur von Anwendungssystemen, Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, leymann@iaas.uni-stuttgart.de
[8] Seven2one Informationssysteme GmbH, Waldstr. 41-43, 76133 Karlsruhe, sven.stando@seven2one.de
[9] Fichtner GmbH & Co. KG, Abteilung für Energiewirtschaft, Sarweystrasse 3, 70191 Stuttgart, Nino.Turek@fichtner.de
[10] Seven2one Informationssysteme GmbH, Waldstr. 41-43, 76133 Karlsruhe, christoph.schlenzig@seven2one.de
[11] Institut für Parallele und Verteilte Systeme, Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, schwarz@ipvs.uni-stuttgart.de

1 Hintergrund und Einleitung

Die Ziele der Deutschen Bundesregierung für die zukünftige Energieversorgung beinhalten weitereichende Veränderungen. Zu installierende Leistungen erneuerbarer Energien sorgen nicht nur für einen Technologiewandel in der Versorgung, sondern auch für eine sich wandelnde Verteilung der Erzeugungseinheiten von wenigen zentralen Einheiten hin zu vielen dezentralen Anlagen und damit neuen Anforderungen an den Netzbetrieb und -ausbau. Ein weiterer Aspekt ist die Dargebotsabhängigkeit der Erneuerbaren Energien, durch die neue Herausforderungen zur Sicherstellung der Systemstabilität und daraus Anforderungen an die weiteren zur Energieversorgung eingesetzten Technologien entstehen.

Um langfristig ein flexibles Energiesystem zu erhalten, in dem erneuerbare Energien Strom erzeugen, welcher bei Überschuss in Batterien, Druckluft- oder Pumpspeichern gespeichert werden kann und die Nachfrage sich flexibel an die aktuelle Erzeugung anpasst, ist jedoch ein Wandel des Verteilnetzbetriebs von Nöten. Dies bedeutet, dass zukünftig noch mehr Technologien im Verteilnetz installiert werden und zusätzlich für ein effektives Lastmanagement eine deutlich detailliertere Informationsbasis gefordert ist. Je mehr die Energiewende auf eine dezentrale Energieversorgung abzielt, desto wichtiger wird der effiziente Betrieb und Ausbau des Gesamtsystems Verteilnetz, inklusive aller sich dort befindlichen Übertragungsleitungen, Erzeugungskapazitäten und Flexibilitätstechnologien. Diese Aufgabe stellt eine deutliche Änderung der bisherigen Anforderungen an den Betrieb der Verteilnetze dar. Daraus stellt sich die Frage, wer diese Anforderungen zukünftig erfüllen wird und welche Handlungsoptionen existieren, um dezentrale Potenziale zur Unterstützung des Gesamtsystems nicht nur technisch, sondern auch wirtschaftlich nutzbar zu machen.

2 Vorgehensweise

In diesem Papier wird ein Konzept eines neuen Marktakteurs vorgestellt (Kapitel 3). In Kapitel 4 wird beschrieben, wie sich dieser im bestehenden Marktumfeld einordnet. Die Handlungsmöglichkeiten des neuen Marktakteurs werden in Abschnitt 5 diskutiert. Daran schließt sich das Konzept für eine optimale Einsatzplanung, dezentrale Betriebsführung und die Realisierung von Flexibilitätsoptionen durch den Akteur an, welches in Abschnitt 6 vorgestellt wird. Allgemeine Anforderungen an die Datenhaltung und an das Asset-Management werden in Abschnitt 7 diskutiert. Eine mögliche Umsetzung mittels Cloud Computing wird in Abschnitt 8 vorgestellt. Abschließend gibt Abschnitt 9 eine Zusammenfassung und einen Ausblick auf das Forschungsvorhaben.

3 Konzept eines neuen Marktakteurs im Verteilnetz – Decentralized Market Agent (DMA)

Vor dem Hintergrund der sich wandelnden Anforderungen im Energiesektor wird eine neue Rolle definiert, deren Aufgabe die Koordinierung von Erzeugern, Speichern und Verbrauchern auf der untersten Netzebene (Verteilnetzebene) sowie dem Verteilnetz als wichtigen zentralen Akteur zur Sicherstellung eines möglichst kostenoptimalen Betriebs und Ausbaus ist. Dazu werden die im Verteilnetz angesiedelten Einheiten durch einen neuen Akteur am Markt vertreten, der durch den Zusammenschluss der Informationen dieses System kostenoptimal und effizient bewirtschaften kann (vgl. Abb. 1). Der Akteur kann zudem vorhandene Produkte an den zentralen Energie- und Regelenergiemärkten im Verteilnetz vermarkten.

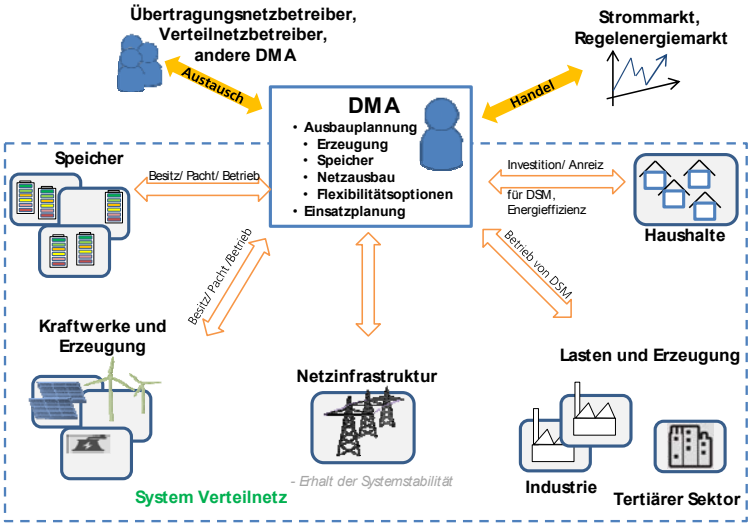


Abb. 1: Konzept des Decentralized Market Agent DMA im Kontext des liberalisierten Strommarktes

Zur Erlösgenerierung dienen folglich der Handel am Spot-Markt sowie den Regelenergiemärkten, aber auch die Bereitstellung von Systemdienstleistungen oder die Generierung vermiedener Netzentgelte. Der Akteur betrachtet neben Erzeugung und Nachfrage auch die Wahrung der Netzstabilität. Durch die Koordinierung der dezentralen Elemente bei gleichzeitigem Handel an den zentralen Marktplätzen wird auch der Zubau von erneuerbaren Energien, Speichern und anderen Flexibilitätskapazitäten, inklusive Netzen, durch diesen Akteur marktgesteuert initiiert. Gleichzeitig erfolgt eine systemdienliche und effiziente Betriebsführung durch den Akteur, der durch den Zusammenschluss bestehender Informationen nicht nur Flexibilitäten im unterjährigen Betrieb besser nutzen kann, sondern auch den Ausbau unter systemischen Aspekten anreizen kann.

4 Bestehende Produkte, Dienstleistungen und Geschäftsmodelle dezentraler Akteure im Strommarkt - Einordnung des DMA im bisherigen Marktumfeld

Im Zuge der Energiewende ergeben sich derzeit mehrere parallele Entwicklungen mit unterschiedlichen, sich teilweise überlagernden Effekten auf die Produkte am Strommarkt sowie die Geschäftsmodelle der Akteure in diesem Umfeld. Wesentliche Treiber hierbei sind:

- Physische Umstellung der Stromerzeugung auf Erneuerbare Energien (EE)
- Zunehmende Dezentralisierung der Stromerzeugung
- Integration der EE in die (bestehenden) Marktmechanismen

Bezüglich der Zielerreichung zur Umstellung der Stromerzeugung auf EE gab es in den letzten Jahren große Fortschritte. So hat sich gegenüber 1998 die installierte Leistung Erneuerbarer Energien mit aktuell rund 85 Gigawatt (GW) mehr als verzehnfacht bei nahezu konstant gebliebener konventioneller Erzeugungsleistung von rund 100 GW [BM15]. Die Aktivitäten der inzwischen mehr als 1 Mio. Stromerzeuger in Deutschland sind jedoch nicht aufeinander abgestimmt und haben jeweils die wirtschaftliche Optimierung in den eigenen Systemgrenzen zum Ziel. Dies führt zu Ineffizienzen im Gesamtsystem. Folgen sind z.B. Netzengpässe, Abregelungen von EE Anlagen oder negative Strompreise.

4.1 Produkte und Geschäftsmodelle

In der folgenden Abbildung wird zusammengefasst, welche Akteure derzeit welche Dienstleistungen in den jeweiligen Marktsegmenten anbieten. Die Marktsegmente gliedern sich in folgende Bereiche Erzeugung dezentral, Strommarkt und Systemdienstleistungen (SDL), Speicher sowie den Bereich der Nachfrage (Stromkunden). Anhand der dargestellten Tätigkeiten kann im nächsten Schritt abgeleitet werden, welche Aufgaben der DMA übernehmen könnte, um eine möglichst effiziente Weiterentwicklung des Gesamtsystems zu unterstützen.

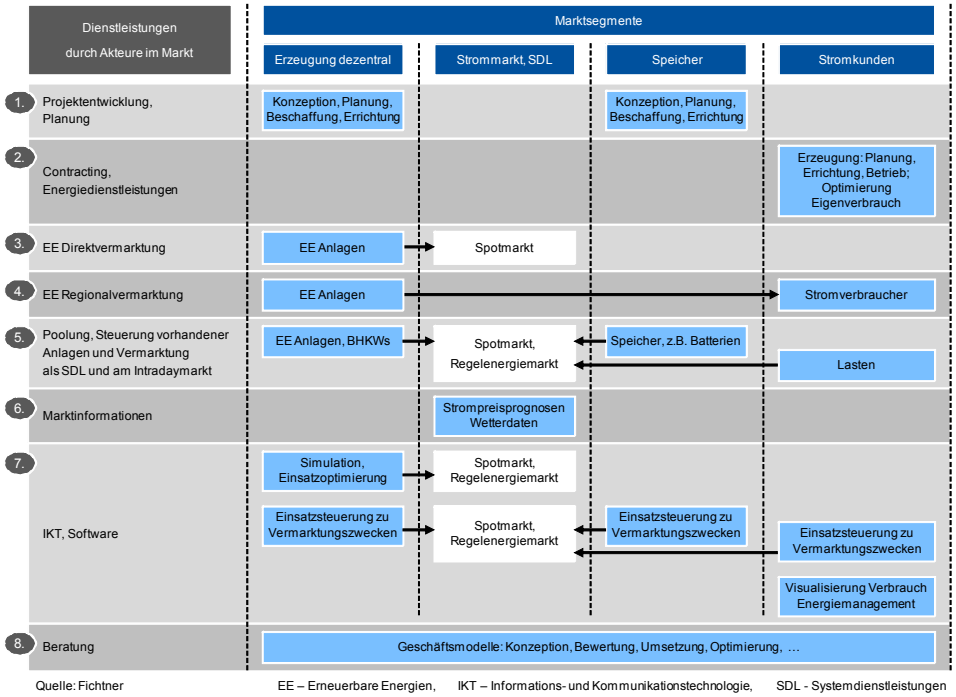


Abb. 2: In Marktsegmenten aktuell erbrachte Dienstleistungen

1. Projektentwicklung und Planung von Anlagen

Die Schritte zur Errichtung von dezentralen Erzeugungsanlagen und Speichern sind die Standortsuche, Konzeption, Planung, Genehmigung, Ausschreibung, Bauüberwachung und Inbetriebnahme der Anlage. Diese Dienstleistungen werden typischerweise von Projektentwicklern oder klassischen Planungsgesellschaften für die Anlagenbesitzer/Anlagenbetreiber erbracht.

2. Contracting, Energiedienstleistungen

Hierunter fallen Planung, Beschaffung, Finanzierung, Errichtung, Betrieb und Wartung einer Anlage durch einen Dienstleister der einem Abnehmer Strom, Wärme, Kälte, Licht, etc. bereitstellt. Der Anlagenbetrieb erfolgt in den Systemgrenzen des Abnehmers. Neu ist ein zunehmender Ausbau dieses Arbeitsgebietes durch Versorger sowie die Erschließung neuer Kundengruppen (z.B. Haushalte) mit neuen Ansätzen und Produkten. Beispiele hierfür sind PV-Pachtmodelle oder Dienstleistungen zur Optimierung von Eigenerzeugung und Energienutzung wie es derzeit z.B. bei der Quartiersversorgung angedacht wird. Eine Herausforderung insbesondere aus wirtschaftlicher Sicht ist die Kleinteiligkeit dieser Aktivitäten, die bei der Bündelung hohe Anforderungen an die technische Infrastruktur setzen.

3. EE Direktvermarktung

Seit Inkrafttreten des EEG 2014 zum 1.8.2014 gilt eine verpflichtende Direktvermarktung für alle Neuanlagen, wenn die installierte Leistung 500 kW übersteigt. Erbracht wird die Dienstleistung von Stromhändlern, Versorgern aber auch von neuen Marktteilnehmern. Die meist von unabhängigen Anbietern aufgebaute Vermarktungsexpertise (einschließlich Börsenzugang) steht jetzt dem Markt (z.B. Erzeugern) als „einkaufbare" Dienstleistung zur Verfügung.

4. EE Regionalvermarktung

Die Vermarktung des aus EE-Anlagen erzeugten Stroms an Abnehmer in regionaler Nähe zu den Anlagen erfolgt vorwiegend durch neue Marktteilnehmer (Direktvermarkter). Die Leistung stellt eine Weiterentwicklung, der mit der Direktvermarktung aufgebauten Expertise und Erschließung neuer Arbeitsgebiete bzw. Angebot neuer Dienstleistungen, dar.

5. Vermarktung vorhandener Anlagen als Systemdienstleistung (SDL) und am Intradaymarkt

a) Vermarktung am Regelenergiemarkt (als SDL)

Physisch erfolgt die kurzfristige Anpassung von Stromerzeugung oder von Lasten (Demand Response), um die Netzfrequenz in der vorgegebenen Bandbreite zu halten. Früher wurden Systemdienstleistungen typischerweise von EVUs meist mit Großkraftwerken erbracht. Regelenergie wird heute wettbewerblich ausgeschrieben und zunehmend kleine Akteure nehmen an dem Regelenergiemarkt teil. Zentrale Treiber für die Veränderungen in diesem Bereich sind die zunehmende Dezentralisierung der Erzeugung, Weiterentwicklungen im Bereich der IKT und regulatorische Anpassungen, die darauf abzielen, mehr Akteuren den Zugang zum Markt zu ermöglichen (z.B. durch Anpassung der Mindestangebotsgröße). Da einzelne, dezentrale Erzeugungsanlagen oder auch Lasten zumeist unter der Mindestangebotsgröße liegen (1 MW für Primärregelleistung und 5 MW für Sekundärregelleistung und Minutenreserveleistung), müssen diese zu „virtuellen Kraftwerken" zusammengefasst („gepoolt") und von einem Akteur gesteuert und vermarktet werden. Neben der Bereitstellung und dem Betrieb flexibler Anlagen erfolgt die zentrale Dienstleistung die Poolung und Steuerung der flexiblen Anlagen entsprechend der Signale der Netzauslastung und der erzielbaren Preise am Regelenergiemarkt.

b) Vermarktung am Intradaymarkt (der Strombörse)

Zur Vermarktung am Intradaymarkt erfolgt eine kurzfristige Anpassung des Betriebs von Erzeugungsanlagen bzw. von Verbrauchern, um auf Preissignale des Marktes zu reagieren. Am Intradaymarkt ist seit 2015 der kontinuierlich Handel von 1/4 - Stundenkontrakten bis zu 45 Minuten vor Lieferung möglich. Soll von dem kontinuierlichen Handel am Intradaymarkt profitiert werden, so kann dies in der Praxis nur mit umfassender IT-basierter Handelsunterstützung gelingen. Für die Vermarkung am Intradaymarkt haben sich unterschiedliche Geschäftsmodelle entwickelt, z.B. die

Bereitstellung entsprechender Softwaresysteme zur Handelsunterstützung, die Bereitstellung von Marktdaten und -prognosen oder die Übernahmen des Handels durch externe Dienstleister.

6. Marktinformationen

Als Dienstleistung werden aktuelle Prognosen (Strompreise, Wetter) typischerweise von speziellen Informationsdienstleistern bereitgestellt. Genutzt werden die Daten um Anlageneinsatz, Beschaffung und Handel (kurzfristig) wirtschaftlich zu optimieren.

7. IKT und Softwareprodukte

Vielen der neuen Geschäftsmodelle liegt die Nutzung von IKT und Softwareprodukten zugrunde. Deren Verfügbarkeit ist Basis unterschiedlicher Dienstleistungsangebote. Die Nutzung von IKT und Softwareprodukten erfolgt in unterschiedlichen Bereichen, z.B.:

- Simulation, Einsatzoptimierung von Anlagen
- Einsatzsteuerung von Anlagen für eine zusätzliche Vermarktung
- Visualisierung von Energieverbräuchen und Energiemanagement

8. Beratung

Aufgrund der massiven Umwälzungen im Strommarkt (Wegbrechen alter Geschäftsmodelle) ist der Bedarf bzw. die Notwendigkeit für die Neuausrichtung etablierter Marktteilnehmer entstanden. Beratungsleistungen werden durch Berater, Ingenieurbüros und Versorger erbracht. Sie können alle zuvor beschriebenen Arbeitsgebiete und Leistungen umfassen. Übergreifendes Ziel ist die Konzipierung, Bewertung, Umsetzung und Optimierung von Geschäftsmodellen.

4.2 Ausblick

Neue Geschäftsmodelle sind insbesondere im Zusammenhang mit der Bereitstellung von Marktzugängen (Spotmarkt bzw. Regelenergiemarkt) und den daraus resultierenden Dienstleistungsmöglichkeiten entstanden. Wie in den vorangegangenen Abschnitten dargestellt, erfolgt die Optimierung heute meist innerhalb der Systemgrenzen der einzelnen Akteure, wodurch Ineffizienzen im Gesamtsystem entstehen. Ursachen sind regulatorische Rahmenbedingungen, nicht abgestimmte individuelle Zielfunktionen (individuelles Optimum vs. Systemoptimum) oder das Fehlen entsprechender (Markt-) Anreize. Um dem zu begegnen, werden im Rahmen des Projektes Ansätze entwickelt, mit denen der DMA einerseits Entwicklungen steuert und andererseits selbst im Markt aktiv wird.

In den Bereichen Projektentwicklung, Planung, Contracting und Energiedienstleistungen ist es die Rolle des DMA, zu steuern und Anreize zu setzen um eine möglichst hohe Effizienz des Systems Verteilnetz zu erreichen. Die EE Direktvermarktung, EE Regionalvermarktung, sowie Poolung und Steuerung vorhandener Anlagen und deren Vermarktung als SDL oder am Intradaymarkt wird der DMA durchführen. Um seine

Aufgaben wahrnehmen zu können, nutzt der DMA Marktinformationen, IKT und Softwareprodukte.

5 Prozesse in einem System „Verteilnetz“ und Handlungsmöglichkeiten des DMA

Bei der Nutzung dezentraler Flexibilitäten ergeben sich unterschiedliche Hemmnisse, die von den bestehenden Akteuren nicht aufgelöst werden. Insbesondere Interessenskonflikte erschweren die Zuordnung von Verantwortlichkeiten aber auch eine systemisch optimale Nutzung der zur Verfügung stehenden Technologien. Die neue Marktrolle DMA soll sich dabei zwischen den bestehenden Akteuren Verteilnetzbetreiber (VNB), Energieversorgungsunternehmen EVU, Betreiber von Einzelanlagen und Endkunden bewegen. Hierzu sind in Abb. 3 die Prozesse und Beziehungen, die in einem System eines Verteilnetzes auftreten und durch den DMA beachtet werden müssen dargestellt. Ein wesentlicher Aspekt aus Sicht des DMA ist das Metering. Die Aggregation der betriebsrelevanten Daten bietet die Möglichkeit möglichst viele Elemente systemeffizient zu nutzen. Zusätzlich kann eine Bündelung von Erzeugungs- und Speicheranlagen für eine verbesserte Handelsposition am Energy-only-Markt oder Regelenergiemarkt erfolgen. In dem Projekt „NEMAR – Netzbewirtschaftung als neue Marktrolle“ werden die einzelnen Prozesse in einem systemanalytischen Energiesystemmodell abgebildet mit dem Ziel, die Möglichkeiten eines Decentralized Market Agents – DMA – dahingehend zu untersuchen, welche Prozesse von diesem Systemplaner übernommen werden können und bei welchen Prozessen die Grenzen des liberalisierten Marktes erreicht oder ggf. überschritten werden.

Um eine umfassende Beurteilung des Konzeptes zu erhalten, werden die Handlungsmöglichkeiten und optimale Ausbauplanung eines Verteilnetzes zudem mittels einer langfristigen Einsatz- und Ausbauoptimierung untersucht. Diese erfolgt mit dem Ziel der Kostenminimierung für das System Verteilnetz. Dies sichert, dass das System in einer wohlfahrtsoptimalen Weise betrieben wird. Durch die Analyse der Modellergebnisse lässt sich feststellen, unter welchen Bedingungen und Preisen die Beziehungen der Akteure im Verteilnetz gestaltet sein müssten, um dieses optimale Ergebnis zu erhalten. Damit kann definiert werden, in welcher Ausgestaltung die Rolle DMA Vorteile für die Bewirtschaftung bringt.

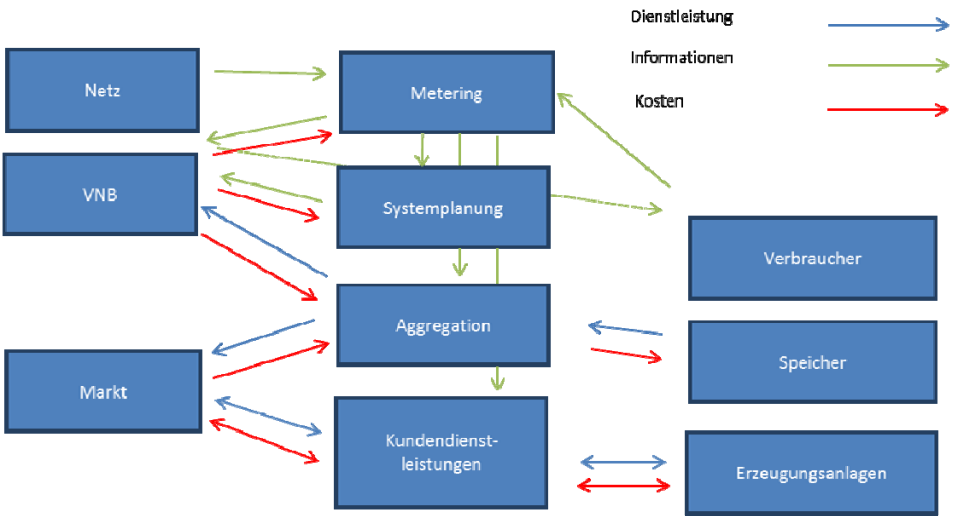


Abb. 3: Skizze unterschiedlicher Prozesse in dem System eines Verteilnetzes und deren Beziehungen untereinander

6 Optimierte Einsatzplanung, dezentrale Betriebsführung und Realisierung von Flexibilitätsoptionen

Die Erschließung potentieller Flexibilitätsoptionen für ein Ensemble von dezentralen Erzeugern, Speichern, Lasten und beeinflussbaren Verbrauchern durch den Akteur DMA erfordert die Vermittlung und Abwägung zwischen den unterschiedlichen technischen und logistischen Anforderungen, welche mit den zu erbringenden Diensten oder Marktprodukten verknüpft sind. Da häufig die dezentralen Komponenten ursächlich für lokale Versorgungsaufgaben zur Verfügung stehen müssen (beispielsweise für die lokale Wärmeversorgung aus dezentralen KWK-Einheiten), können sich hierdurch strenge Einschränkungen an die optimierte Betriebsführung ergeben, welche der DMA bei der Einsatzplanung und Vermarktung berücksichtigen muss. Insbesondere bei multivalenten Systemen, bei denen neben Strom auch noch andere Energieprodukte wie Wärme- oder Kälteversorgung involviert sind, hängen die lokalen Anforderungen zudem auch noch von äußeren Parametern wie beispielsweise den meteorologischen Bedingungen oder den variierenden Nutzungsprofilen unterschiedlicher Gebäudetypen ab. Daher ist eine sorgsame Analyse der lokalen Randbedingungen und Anforderungen für dezentrale Flexibilitäten ein wichtiger Ausgangspunkt für die Einsatzplanung durch den DMA.

Die Vorgehensweise zur Integration der dezentralen Flexibilitäten in die übergeordnete Handlungsebene des DMA vollzieht sich in einem dreistufigen Prozess:

(1) Einsatz- und Zustandsbewertung der Einzeleinheiten zur Erstellung einer

quantifizierten Flexibilitätsprognose

(2) Gesamteinsatzoptimierung durch Auswahl aus allen Flexibilitätsoptionen unter Beachtung der jeweiligen lokalen Rahmenvorgaben durch den DMA, Generierung von Einsatzfahrplänen

(3) Implementierung der Einsatzfahrpläne im Betrieb und notwendige laufende Anpassungen bei Fahrplanänderungen und -abweichungen.

Im ersten Schritt werden neben den lokalen Anforderungen an die Betriebsführung bestehende Anlagenflexibilitäten qualitativ und quantitativ charakterisiert, wobei sowohl technische als auch ökonomische Bewertungsparameter beschrieben werden. Dies bedeutet beispielsweise im Fall elektrischer Raumheizsysteme, neben Heizleistung und thermischen Anforderungsprofilen auch thermische Gebäudekapazitäten sowie Kosten alternativer Heizungsoptionen zu evaluieren. Ergebnis wären in diesem Beispiel Zeitreihen von verschiebbarer oder beeinflussbarer elektrischer Verbrauchsleistung sowie mögliche Kosten, die mit dem Abruf dieser Flexibilität verbunden sind. Im Fall dezentraler Erzeuger (wie Photovoltaik) spielt als Betriebsführungsziel insbesondere auch die Erreichung eines hohen Eigenversorgungsgrades eine besondere Rolle, wobei im gegebenen Fall diese Anforderung die Nutzung anderer Flexibilitätsoptionen einschränken kann.

Anschließend vermittelt der DMA die dezentralen Flexibilitäten mit Dienstleistungsoptionen und Marktprodukten, um eine integrative, technisch und ökonomisch optimierte Leistungserbringung für Kunden oder Märkte zu erzielen. Der DMA Ansatz erlaubt dabei eine Kombination verschiedener Funktionalitäten (wie die gleichzeitig Teilnahme am Strommarkt und die Bereitstellung von Regelenergie), was zu hoher technischer und ökonomischer Effizienz führt. Nach erfolgter Vermarktung der Gesamtflexibilität ist es dann Aufgabe des DMA, die Erbringung der technischen Leistungen auf die Einzeleinheiten aufzuteilen und für diese Anforderungsfahrpläne zu erstellen. Zu beachten ist dabei, dass bei einer Reihe von möglichen Dienstleistungen (Regelenergie, Bilanzkreismanagement, Netzstützung) die eigentliche Inanspruchnahme der Dienstleistung unprognostizierbar per Abruf erfolgt, was zu unvermeidbaren Abweichungen von den ursächlichen Fahrplanvorgaben führt. Hier ist bereits in dieser Phase der Einsatzplanung durch entsprechende Planungsreserven sicherzustellen, dass auch im „worst case“ das Ensemble der dezentralen Anlagen allen Flexibilitätsanforderungen gerecht werden kann.

Im dritten Schritt müssen die Anforderungsfahrpläne des DMA in reale Einsatzfahrpläne für die Einzelkomponenten umgesetzt werden, wobei zum Zweck der Komplexitätsreduktion dieser anlagennahe Schritt durch dezentrale Managementsysteme realisiert werden kann, die sich am Installationsort der Anlagen befinden. So wird beispielweise bei Anforderung einer Leistungsänderung eines Ensembles aus 3 BHKW Aggregaten auf Basis einer lokalen Einsatzoptimierung die Entscheidung zur Auswahl des geeignetsten der Aggregate getroffen und werden Steuerungsbefehle für lokale SCADA Systeme generiert. Die Anwendung

modellprädiktiver Regelungsverfahren erlaubt hierbei eine fortlaufende Anpassung der lokal optimierten Einsatzfahrpläne. Gleichzeitig muss sichergestellt werden, dass die dezentralen Einheiten dem DMA fortlaufend aggregierte Betriebs- und Zustandsdaten zur Verfügung stellen, die diesem die Überwachung und ggf. Beeinflussung der Gesamtbetriebsführung aller beteiligten Flexibilitäten ermöglicht.

7 Datenhaltung und Asset-Management in der Energiewirtschaft

Auch wenn die Rolle des Decentralized Market Agent (DMA) noch nicht vollumfänglich ausgestaltet werden kann sind zentrale Anforderungen an diesen bereits eindeutig ableitbar. Um als dezentraler Marktakteur den Betrieb und Ausbau des Verteilnetzes kostenoptimal gewährleisten zu können, benötigt dieser eine Vielzahl an Informationen. Informationen zu den einzelnen Teilnehmern im Netzgebiet, den Anlagen und deren verfügbaren Flexibilität gilt es genauso zu verwalten wie relevante externe Daten. Dazu können Marktpreise an den Spot oder Regelenergiemärkten, aber auch Prognosen zu Wetter und Last gezählt werden. Sowohl Stammdaten als Zeitreihendaten spielen für den kostenoptimalen Betrieb eine entscheidende Rolle um die Verfügbarkeiten des Kraftwerksparks zu berechnen und dessen Flexibilitäten erkennen zu können und vermarktbar zu machen.

Die Anforderungen an die Daten und die entsprechende Datenhaltung unterscheidet sich von derer der traditionellen Energiewirtschaft vor allem insofern, dass die einzelnen Teilnehmer, Anlagen und weiteren relevanten Faktoren noch nicht eindeutig identifizierbar sind und sich auch über die Zeit deutlich schneller verändern. Das hat zur Folge, dass ein System zur Abbildung dieser Daten über eine möglichst hohe Flexibilität und Erweiterbarkeit verfügen muss um schnell und effizient auf Änderungen reagieren zu können.

Der Aufbau und Unterhalt von IKT-Systemen zur Verwirklichung solcher neuer Geschäftsmodelle ist oft mit prohibitiv hohen Kosten verbunden, die den wirtschaftlichen Erfolg einer Geschäftsidee gefährden können. Zudem ist das zukünftige Marktumfeld eines DMA heute noch recht ungewiss. Unklar ist, welche regulatorischen Rahmenbedingungen gelten werden oder an welchen Märkten Flexibilitäten zukünftig handelbar sind.

8 Einsatz von flexiblen und dem Bedarf ausgerichteten Cloud Architekturen als Schnittstelle zwischen dem DMA, der Messinfrastruktur und der Datenhaltung

Für die kostenoptimale Regelung von Verteilnetzen müssen Daten von verschiedenen Quellen, wie Blockheizkraftwerken, Speichern und Haushalten aus dem Verteilnetz zusammengeführt und ausgewertet werden. Dafür bedarf es einer Plattform, die es

erlaubt, Datenquellen flexibel und nach Bedarf anzubinden, um sowohl der Umstrukturierung und dem Ausbau von Verteilnetzen und den damit einhergehenden neuen Datenquellen, als auch der immer stärkeren Ausweitung des Einsatzes von Smart-Metering-Lösungen Rechnung zu tragen. Des Weiteren müssen neben den erforderlichen Auswertungs- und Analysealgorithmen zur kurzfristigen und oft ausgeführten kostenoptimalen Regelung des Verteilnetzes auch Algorithmen zur Generierung von längerfristigen Ausbauplanungen in größeren Zeitabständen auf den in der Plattform zusammengeführten Daten operieren können. Daraus lassen sich im Wesentlichen zwei Workloadszenarien ableiten, welche die Plattform zu handhaben im Stande sein muss. Die Plattform muss entworfen werden, um zum einen kontinuierlich steigende Workloads durch die wachsende Anzahl an Datenquellen für die kurzfristigen Fahrplangenerierungen verarbeiten und zum anderen periodisch auftretende Workloadspitzen für die Ausbauplanung abfangen zu können. Da ein DMA auf der Ebene eines Verteilnetzes agiert, sollte es zudem möglich sein, die Plattform sukzessive für eine steigende Anzahl von Verteilnetzen bereitzustellen. Dabei muss die technische Komplexität des Systems für den DMA so transparent wie möglich sein, damit die Bereitstellung der Plattform in weiteren Verteilnetzen nicht zum großen Kostenfaktor in Verbindung mit hohen Anschaffungs- und Einrichtungskosten von neuer Hard- und Software wird.

Dementsprechend muss der DMA durch ein Softwaresystem unterstützt werden, welches vor allem hinsichtlich der (i) entstehenden Kosten, (ii) flexiblen Anbindung von Datenquellen, (iii) über die Zeit kontinuierlich steigenden Workloads aufgrund einer steigenden Anzahl an Datenquellen, (iv) periodisch auftretenden Workloadspitzen und (v) durch effiziente Provisionierung neuer Instanzen eine zukunftsorientierte Lösung darstellt.

Für eine Rolle wie den DMA existiert bislang noch keine fertige Software-Lösung, deshalb muss das System für den DMA neu entwickelt werden. Cloud Computing, als ein neues Paradigma der Nutzung und des Anbietens von IT-Ressourcen, stellt hierfür eine Möglichkeit dar. Solche Ressourcen können bedarfsweise von jedem, der Zugriff auf das Internet hat, benutzt werden. Ressourcen werden nutzungsbasiert abgerechnet [Ley09]. Der Hauptgrund Cloud Computing zu verwenden ist, vom Modell der Investitionskosten zum Modell der Betriebskosten zu wechseln: Anstatt IT-Ressourcen wie Maschinen, Speichereinheiten oder Software zu kaufen und Personal für den Betrieb oder die Wartung anzustellen, bezahlt eine Firma eine andere (den Cloud-Anbieter) für die tatsächlich verwendeten Ressourcen (pay-as-you-go). Dies verspricht eine Kostensenkung entsprechend Anforderung (i) [VHH12]. Ein wesentliches Merkmal ist die Elastizität. Elastizität bedeutet, dass eine Ressource dem Benutzer immer zur Verfügung steht und dass diese bedarfsweise wächst oder schrumpft. Als grundsätzliche Unterscheidungsmerkmale von Diensten, die von der Cloud angeboten werden, haben sich drei Kategorien herauskristallisiert: Software as a Service (SaaS), Platform as a Service (PaaS) und Infrastructure as a Service (IaaS) [MG11]. Im Falle von IaaS wird Hardware virtualisiert und der Softwareanbieter muss die zum Betrieb seiner Software erforderlichen Komponenten selbst auf virtuellen Maschinen bereitstellen. Bei SaaS wird

eine Softwarelösung dagegen als Gesamtpaket angeboten, die als Ganzes benutzt und deren Komponenten transparent vom Cloud-Anbieter skaliert werden. Im Falle von PaaS werden Softwarekomponenten, wie Datenbanken oder Message-Queue-Systeme, von dem Cloud-Anbieter angeboten und nach Bedarf skaliert. PaaS-Komponenten bieten sich insbesondere für den Umgang mit kontinuierlich steigenden Workloads und periodisch auftretenden Workloadspitzen an, da diese elastisch sind und somit auf sich ändernden Workload reagieren können (Anforderungen iii und iv). Für die in Anforderung (ii) beschriebene flexible Anbindung von Datenquellen muss eine Kopplungs- und Speicherungsart gewählt werden, die mit heterogenen Datenformaten umgehen kann. Hierfür bieten sich beispielsweise die PaaS-Lösungen Message-Queueing-Systeme (z.B. Amazon Simple Queue Service [Am15a]) und Key-Value-Stores (wie z.B. Amazon SimpleDB [Am15b]) an. Wird eine Anwendung neu entwickelt, so wird hierbei auf bewährte Lösungsbausteine, sogenannte Patterns, zurückgegriffen. Im Falle von Cloud-basierten Anwendungen wurden diese von Fehling et al. [Fe14] dokumentiert. Auf Basis der Patterns werden dann Softwarekomponenten neu entwickelt, bestehende angepasst und SaaS- sowie PaaS-Angebote von Cloud-Anbietern ausgewählt und in die Anwendung integriert [An14]. Zur Provisionierung und Skalierung der gesamten Anwendung, unabhängig von den gewählten Cloud-Angeboten, eignet sich der TOSCA-Standard [Bi12] und unterstützende Werkzeuge. Er erlaubt eine standardbasierte Beschreibung der Topologie der Anwendung und ermöglicht TOSCA-konformen Laufzeitumgebungen, wie OpenTOSCA [Bi13], diese zu betreiben. Das entwickelte DMA-Gesamtsystem wird folglich in einem TOSCA-Paket paketiert um eine effiziente Bereitstellung nach Bedarf zu ermöglichen. Für jedes neue Verteilnetz, in dem der DMA agieren soll, kann von einer TOSCA-Laufzeitumgebung eine neue Instanz erstellt und verwaltet (Anforderung v) werden.

9 Zusammenfassung und Ausblick

Ziel des in diesem Paper vorgestellten Projektes NEMAR ist die Entwicklung eines Mechanismus, dem es gelingt, dezentrale Potenziale an Flexibilitäten des Stromsystems zur Unterstützung des zentralen Netzverbundes zu mobilisieren. Dieser zu entwickelnde Mechanismus soll durch eine neue Marktrolle ausgeführt werden, deren Aufgabe die effiziente Bündelung der Flexibilitäten und deren optimierter dezentraler Einsatz und Ausbau ist. Die Marktrolle dient der ökonomisch effizienten Betriebs- und Ausbauplanung unter Aufrechterhaltung der Systemstabilität. Dabei richtet sie sich nach den Signalen der bestehenden zentralen Marktplätze.

Um diesen neuen Marktakteur zur Ausfüllung des Konzeptes zu befähigen, bedarf es weitreichender Untersuchungen und Entwicklungen. Als Grundlage für die kommenden Arbeiten wurden in diesem Paper derzeit bestehende Produkte, Dienstleistungen und Geschäftsmodelle dezentraler Akteure im Strommarkt (Kapitel 3) beschrieben. Das System Verteilnetz wurde zudem hinsichtlich unterschiedlicher Prozesse untersucht (Kapitel 4) und eine dezentrale Regelung und Einsatzoptimierung der einzelnen

Elemente in diesem System exemplarisch dargestellt (Kapitel 5). Um die Rolle des DMA umsetzen zu können benötigt der Akteur eine umfassende, flexible IKT-Infrastruktur, die die Aggregation und Aufbereitung der Daten ermöglicht. Hierzu wurde in Kapitel 6 die Möglichkeit der Datenhaltung und des Asset-Managements beleuchtet und darauf folgend eine flexible Cloud Architektur vorgestellt, die als Schnittstelle zwischen dem DMA, der Messinfrastruktur und der Datenhaltung eingesetzt werden kann (Kapitel 7).

Insgesamt zeigt sich, dass für die zukünftige Nutzung „smarter“ Infrastrukturen in der Stromversorgung weitreichender Forschungsbedarf besteht. Die konzeptionelle Entwicklung und im Projekt „NEMAR“ folgende Evaluierung des Decentralized Market Agent – DMA – soll einen Beitrag leisten, Möglichkeiten zur Hebung dezentraler Flexibilitäten zu entwickeln, um diese dem zentralen Netzverbund zur Verfügung zu stellen. Hierzu wurde in diesem Paper das Konzept des DMA grob erläutert. Hierbei wurde angesprochen, welche Aspekte nötig sind, um einen DMA zu installieren und welche regulatorischen Implikationen durch die Einführung einer solchen Rolle DMA entstehen können. Die detaillierte Analyse und Bewertung der Chancen aber auch die Grenzen des DMAs im Rahmen des liberalisierten Marktes sind ein wesentlicher Teil der Untersuchungen im vorgestellten Projekt „NEMAR“.

Danksagung: Die Autoren danken dem BMWi für die Förderung im Rahmen des Projekts NEMAR (03ET4018).

Literaturverzeichnis

[AB00] Abel, K.; Bibel, U.: Formatierungsrichtlinien für Tagungsbände. Format-Verlag, Bonn, 2000.

[ABC01] Abraham, N.; Bibel, U.: Corleone, P.: Formatting Contributions for Proceedings. In (Glück, H.I. Hrsg.): Proc. 7th Int. Conf. on Formatting of Workshop-Proceedings, New York 1999. Noah & Sons, San Francisco, S. 46-53, 2001.

[Am15a] Amazon. Simple Queue Service (SQS). http://aws.amazon.com/de/sqs/, 2015.

[Am15b] Amazon. SimpleDB. http://aws.amazon.com/de/simpledb/, 2015.

[An14] Andrikopoulos, V.; Darsow, A.; Karastoyanova, D.; Leymann, F.: CloudDSF - The Cloud Decision Support Framework for Application Migration. In: Proceedings of the Third European Conference on Service-Oriented and Cloud Computing (ESOCC 2014), S. 1-16, Springer Berlin Heidelberg, 2014

[Az09] Azubi, L. et.al.: Die Fußnote in LNI-Bänden. In (Glück, H. I., Hrsg.): Formatierung 2009. LNI 999, Format-Verlag, Bonn, S. 135-162, 2009.

[Bi12] Binz, T.; Breiter, G.; Leymann, F.; Spatzier, T.: Portable Cloud Services Using TOSCA. In: IEEE Internet Computing. Vol. 16(03), IEEE Computer Society, 2012.

[Bi13] Binz, T.; Breitenbücher, U.; Haupt, F.; Kopp, O.; Leymann, F.; Nowak, A.; Wagner,

S.: OpenTOSCA - A Runtime for TOSCA-based Cloud Applications. In: Proceedings of 11th International Conference on Service-Oriented Computing, 2013.

[Ez10] Ezgarani, O.: The Magic Format – Your Way to Pretty Books. Noah & Sons, 2010.

[Fe14] Fehling, C.; Leymann, F.; Retter, R.; Schupeck, W.; Arbitter, P.: Cloud Computing Patterns, Springer, 2014.

[GI14] GI, Gesellschaft für Informatik e.V., www.gi-ev.at, Stand: 24.12.2014.

[Gl09] Glück, H. I.: Formatierung leicht gemacht. Formatierungsjournal 11/09, S. 23-27, 2009.

[MG11] Mell, P.; Grance, T.: The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology, Special Publication 800-145, 2011.

[Ley09] Leymann, F.: Cloud Computing: The Next Revolution in IT. In (Fritsch, D., Hrsg.): Photogrammetric Week ’09. Wichmann Verlag, S. 3-12, 2009.

[VHH12] Vossen, G.; Haselmann, T.; Hoeren, T.: Cloud Computing für Unternehmen. dpunkt.verlag, 2012

[Wa14a] Wasser, K.; Feuer, H.; Erde, R.; Licht, H.: Essenzen der Informatik. Verlag Formvoll, 2014.

[Wa14b] Wasser, K.; Feuer, H.; Erde, R.; Licht, H.: Ganz neue Essenzen der Informatik im selben Jahr. Format-Verlag, 2014.

[BM15] Bundesministerium für Wirtschaft und Technologie (BMWi) (Hg.) (2015): Energiedaten. AG Energiebilanzen e.V. Online verfügbar unter http://www.bmwi.de/DE/Themen/Energie/Energiedaten-und-analysen/Energiedaten/gesamtausgabe,did=476134.html, zuletzt geprüft am 13.05.2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Towards a Cloud-based Platform Architecture for a Decentralized Market Agent

Oliver Kopp[1] Michael Falkenthal[2] Niklas Hartmann[3] Frank Leymann[4] Holger Schwarz[5] Jessica Thomsen[6]

Abstract: Reorganization of power generation, thereby replacing conventional energy sources by innovative renewable energy sources, demands a change in distribution grid structure and operation. The foreseen Decentralized Marked Agent is a new role in the energy market sector accomplishing not only trading on energy and operating reserve markets but also regulating flexibilities at the distribution grid level, such as energy storage and decentralized energy generators, and thereby considering system services and securing system stability. This paper presents requirements on an IT system to support this new role. We design an architecture matching these requirements and show how Cloud computing technology can be used to implement the architecture. This enables data concerning the distribution grid being automatically gathered and processed by dedicated algorithms, aiming to optimize cost efficient operation and the development of the distribution grid.

1 Introduction

To match the objectives of the German energy transition [BM10], excessive changes will be necessary. Shifting towards power plants based on renewable energy sources consequently not only demands predominantly innovative technologies but also a change in the energy supply structure. That means, the current centralized energy supply structure composed of large central power plants must be replaced by a structure consisting of more decentralized power plants. Thus, the distribution grid structure, operation, and expansion have to adapt to new energy generation patterns. Considering these arising or already existing challenges, the new market role "Decentralized Market Agent" (DMA) is envisioned in the energy sector [Th15]. The DMA addresses problems in coordination of the synergy of energy generation, storage systems, and variety of the consumers on distribution grid level. Additionally, in order to ensure a cost efficient operation and expansion, the distribution grid itself is an important factor in future energy systems. To realize this task, the DMA has to be efficiently

[1] Institute for Parallel and Distributed Systems, Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, Germany, kopp@ipvs.uni-stuttgart.de
[2] Institute of Architecture of Application Systems, Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, Germany, falkenthal@iaas.uni-stuttgart.de
[3] Fraunhofer-Institut für Solare Energiesysteme ISE, Energy Systems and Markets, Heidenhofstraße 2, 79110 Freiburg, Germany, niklas.hartmann@ise.fraunhofer.de
[4] Institute of Architecture of Application Systems, Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, Germany, leymann@iaas.uni-stuttgart.de
[5] Institute for Parallel and Distributed Systems, Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, Germany, schwarz@ipvs.uni-stuttgart.de
[6] Fraunhofer-Institut für Solare Energiesysteme ISE, Energy Systems and Markets, Heidenhofstraße 2, 79110 Freiburg, Germany, jessica.thomsen@ise.fraunhofer.de

supported by IT. During the past years, Cloud computing established itself as an IT cost saving alternative and enabled companies to focus on their core business [VHH12]. Thus, Cloud computing seems to be a suitable basis especially due to the associated paradigm of rapid provisioning of IT resources coupled with expenditures for only actually used resources. Based on these facts, we envision a Cloud-based platform enabling the DMA to fulfill its dedicated task. The Cloud-based platform is realized by an IT system that is aligned and optimized to the needs of the DMA.

In this paper, we exemplify the DMA and its environment in Sect. 2. In Sect. 3 we abstract from the DMA environment and identify the core components and concepts, which have to be supported by an IT system. Based on that, we derive requirements on a platform in Sect. 4 and propose a platform fulfilling these requirements in Sect. 5. The Cloud-readiness of that platform is discussed in Sect. 6. Related work is the focus of Sect. 7. Finally, we conclude the paper and provide an outlook on future work in Sect. 8 .

2 Overview on the DMA

With regard to the energy transition the DMA aims to incorporate all elements located on distribution grid level ensuring cost efficient operation. Thus, the DMA represents the aggregated entities at the central markets and will not only trade on energy-only and reserve markets, but will also consider system services and secure system stability. By

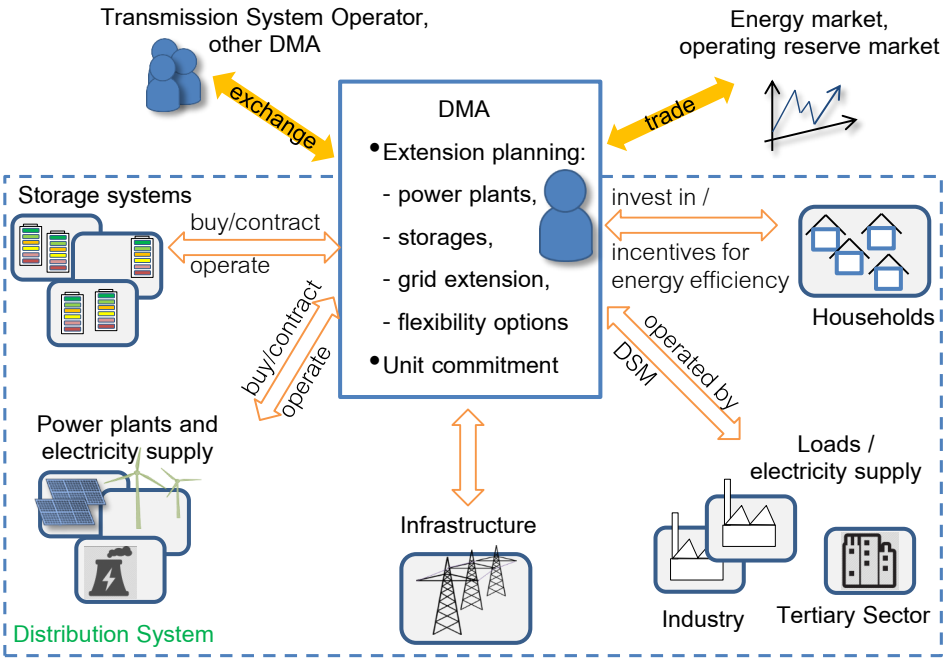


Fig. 1: The DMA in the liberalized electricity system [Th15]

coordinating decentralized elements while simultaneously trading at the central markets, the expansion of energy generation, storage, and other flexibility capacities can be stimulated based on market incentives. At the same time, the DMA is responsible for an efficient system operation. The new actor can benefit from aggregating existing information and, hence, from employing more flexibility options within the annual operation. Furthermore, aggregating information from generation, consumption, and grid operation additionally enables a more system friendly expansion of each distribution grid system. The embedding of the DMA in the liberalized electricity system is presented in Fig. 1. In general it becomes apparent that the DMA focuses on two major activities: First to optimize a distribution grid in short term schedules, i. e., in an interval of minutes, in order to regulate its actual utilization, and second to optimize the constructional development and improvement of the grid's topology and its physical components. The latter optimization scenario envisioned to be performed several times a year.

3 Abstracted Environment of the DMA

Mathematical approaches are the basis for short term as well as for long term optimization. The approaches rely on models of the distribution grid that are optimized based on specifically defined optimization criteria. These criteria are specific for each instance of a DMA, i. e., for each distribution grid, because the models have to take into account the local conditions of the distribution grid, which is managed by the DMA. One optimization criteria could be to achieve the cost optimum for the whole distribution grid while strictly avoiding any disadvantageous situations that could lead to blackouts of the grid. In this scenario especially critical factors such as weather conditions have to be considered as they have strong effects on the power generation by renewable energies. Another optimization criteria would be to minimize the emission of green house gases. In this scenario the major goal consists of avoiding fossil fuels and covering as much power generation as possible using renewable energy sources. In real world scenarios, however, a combination of several optimization criteria being weighted against each other seems to be feasible to leverage the DMA's needs.

In principle sensors and actuators form the two major components the DMA has to deal with. Sensors are various meters, brought out in the distribution grid acting as data sources for the DMA. Meters are mainly proprietary devices collecting data in heterogeneous formats, intervals, and details. Actuators on the contrary are flexibility points in the distribution grid that can be regulated by the DMA. They might be points to manipulate the power generation within the distribution grid, such like block-type thermal power stations or wind power plants. Additional points might be flexibilities regulating power consumption and storages in the form of batteries and pumped storage hydro power stations. Finally, agents interacting on energy markets, where energy can be traded and sold to nearby distribution grids, are realizations of the actuator concept. Thus, the concept of an actuator subsumes every logic and control unit of the distribution grid and all other regulation options to influence the distribution grid.

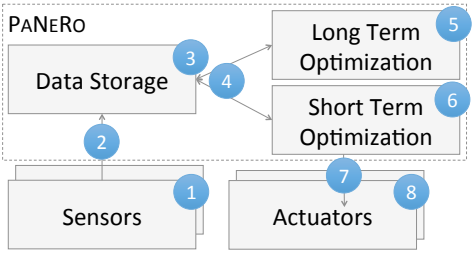


Fig. 2: Components of the DMA environment

The algorithms implementing the short term and the long term optimization operate on historic data about the distribution grid. Therefore, it is important to meter data from sensors and to store them over time. Thus, the platform of the DMA can conceptually be described as a data storage component and two optimization components. The data storage collects metering data from the sensors whereas the two optimization components perform the long term and the short term optimizations. Based on the outcomes of the short term optimization algorithm the actuators are regulated. We call the platform of the DMA PANERO: **Pla**tform for the **Ne**w market **Ro**le. It is illustrated in Fig. 2, which additionally shows eight variation points being discussed in the next section where we discuss requirements on PANERO.

4 Requirements on the Platform Architecture

The main purpose of a DMA is both to manage energy supply at the level of a distribution grid and to develop and extend a distribution grid based on optimization criteria. To automate data collection and analysis, an IT system is required. In total eight variation points have been identified and depicted as numbered circles in Fig. 2. They emphasize specific requirements to PANERO. The requirements have been derived during discussions in the project and are based on the expertise of the involved cooperation partners from the fields of distribution grids, power plant construction, data analysis, database technologies, and Cloud application architecture. This paper presents the current state of the discussion. Currently, concrete requirements on time-constraints are missing. That means, the maximum time, the data store may take to process a query and the maximum time the short term optimization algorithm may take to create commands for the actuators is not fixed yet.

Below, elicited requirements are consecutively stated for all eight variation points. The range of feasible realization options are discussed for each point to show the complex space of possible solutions, which, of course, lead to a vast number of possible concrete realizations of PANERO. Nevertheless, the summary of requirements and possible solutions (see Sect. 5) document major design decisions to make in order to architect and develop an IT system for the DMA. The stated requirements and solutions are not only applicable to develop PANERO. Since PANERO is only one realization of an IT system to support the management of a distribution grid the raised requirements and solutions can conceptually be used as a basis for the development of a whole archetype of IT systems.

Requirement 1: Support heterogeneous sensors

The sensors (variation point 1) may offer their data using a representational state transfer (REST) or Web Service endpoint, or by protocols such as IEC 61850 [IE13] or IEC 62056-21 [IE02], which are both standards in the domain of digital control systems. It must be possible to connect all these sensors to the platform to obtain the values (R1).

Requirement 2: Secure, loosely coupled, and reliable connection to the sensors

The connection of the sensors to the platform (variation point 2) should be secure (R2a). Security is required to ensure privacy and avoid manipulation by third parties [Fe14a].

The connection of the sensors to the platform should be loosely coupled (R2b). Typically, loose coupling means autonomy in the cases of platform, reference, time, and format [Fe14b]. *Platform autonomy* states that differences in platforms, e. g., regarding data representation or processor architecture, are unimportant for the communication. That is required to enable the sensors and PANERO to be implemented on different platforms. *Reference autonomy* states that the concrete network address of an endpoint should not be known. That is required to enable the platform to be scaled when new sensors are deployed during runtime. *Time autonomy* states that the receiver has not to be online when the sender is sending data. Also, communication partners may send and receive information at their own speed, thus, if one communication partner transmits information faster than another one, both remain operational. *Format autonomy* states that the data format used for exchange should be understood by both the sender and receiver. That format may be different from the format used internally.

Since some sensors can provide data about critical infrastructures of a distribution grid, those have to be reliably transmitted to the platform. Reliable transfer causes overhead in the communication. Therefore, it is enough to support reliable transfer only for sensors of which the values are critical for the optimization algorithms. As a result, the connection has to be reliable for specific sensors (R2c).

Requirement 3: Data store scaling and triggering capabilities

The data store (variation point 3) has to scale with the connected sensors (R3a). That means, if more sensors are connected or the rate of sensor data is increased, the data store has to be able to handle these increasing loads in terms of storage capacity and the number of input and output operations.

The platform has to be able to trigger the short term optimization in the case exceptional data arrives (R3b). Examples for exceptional data are values differing by a certain threshold from other values or values indicating an outage of a generation unit.

Requirement 4: Suitable data format and support for time series

The data store has to offer (variation point 4) the data in a format required by the optimization algorithms (R4a). This is required as it is assumed that the data store can convert the data more efficiently than the optimization algorithms. Further, this requirement enables to introduce a separate of concerns by separating data preparation logic from the optimization logic of the algorithms. Thus, data preparation logic does not pollute the optimization algorithms so that developers can concentrate on the actual valuable logic of the algorithms.

Further, one sensor date is always related to one specific timestamp, thus, sensor data has in general to be handled as time series. Since data connections to sensors are not guaranteed to be reliable in some cases (see variation point 2) and data sampling rates differ, the data store should provide functionality to interpolate data. As a result the data store should offer time series functionality such as selection of data ranges, aggregation of values, or interpolation of missing data (R4b). Alternatively, the optimization algorithms can do the interpolation by themselves. It is, however, assumed that the data store will perform these operations more efficient than the optimization algorithms. Providing such time series functionality natively by the platform also enables benefits by means of separation of concerns regarding the optimization algorithms as already mentioned above for the preparation of appropriate data formats, accordingly.

Requirement 5: Support for function shipping

Function shipping versus data shipping is a discussion for gaining performance speed-up. Function shipping describes that functions, which in case of PANERO contain logic of optimization algorithms, are shipped to parts of a system where they can optimally run close to the data in order to avoid tedious data transfers [Or00]. By contrast, data shipping describes the process that data is queried by a component and shipped to the component for processing. To achieve performance improvements regarding the optimization components (variation points 5 and 6), it has to be possible to ship the implemented optimization algorithms to the database (R5).

Requirement 6: Secure, reliable, loosely-coupled, and time-constraint connection to the actuators

The commands sent to the actuators should not be viewable or changeable by third parties. Therefore, the connection to the actuators (variation point 7) has to be secure (R6a).

It has to be assured that the commands arrive at the actuators. Therefore, connection and data transfer to the actuators have to be reliable (R6b).

The message format has to be generic to enable loose coupling between the platform and the actuators (R6c). The same arguments as presented in the context of R2b apply.

The commands sent to the actuators have to arrive within a certain time frame. Therefore, the connection to the actuators has to offer delivery guarantees with respect to the maximum delivery time (R6d).

Requirement 7: Plausibility checks

The actuators (variation point 8) may evolve during the runtime of the optimization algorithm. Therefore, the actuators have to be enabled to do a plausibility check of the received commands to avoid implausible regulations (R7). Regarding energy generation and storage systems this means that actuators guarantee operation within predefined thresholds to prevent failures.

5 Proposed Platform Architecture

Fig. 3 presents an architecture satisfying the aforementioned requirements. The general idea of this architecture is to introduce (i) data gateways serving as an abstraction layer to cover the complexity of sensors and actuators and (ii) to use message queuing to enable specific qualities of service regarding the connections of the sensors and actuators to the platform. The data gateway is responsible to transform the data gathered by the sensors to an intermediate format. It is also used to transform the commands received by actuators to their proprietary format. Furthermore, message queuing is a popular technique to implement loose coupling [HW03]. In the following, we discuss how the proposed architecture satisfies the requirements.

To support heterogeneous sensors (R1), a data gateway is introduced. It is a component forming the gateway between concrete sensors as well as actuators and PANERO. In case the sensors offer their data using a REST endpoint, they can be directly connected to a message queue. In case the sensors offer their data using protocols such as IEC 61850, they can be connected using the OpenMUC platform [Op15], which is an open source software framework for monitoring and control applications in smart grids, and a custom plugin for OpenMUC connecting to the message queue.

To satisfy R2b (loose coupling), a message queuing system (MQS) [HW03] is introduced. The queues managed by the MQS are shown in the figures as pipes. By using a MQS, all the properties of loose coupling can be achieved. Platform autonomy is achieved, since MQS implementations are available for different platforms and can communicate with each other. Reference autonomy is achieved, because the sender puts its message in its local queue and the MQS takes care about the routing. Time autonomy is achieved, because the sender has to put the message in its local queue and the MQS takes care about forwarding it to the queue of the receiver. This way sender and receiver are decoupled since the receiver can process messages from its queue one by one and independently from any sender. Format autonomy is achieved, because MQS allow to convert messages to any appropriate data format during transfer if required. When choosing a concrete MQS, the system has to offer

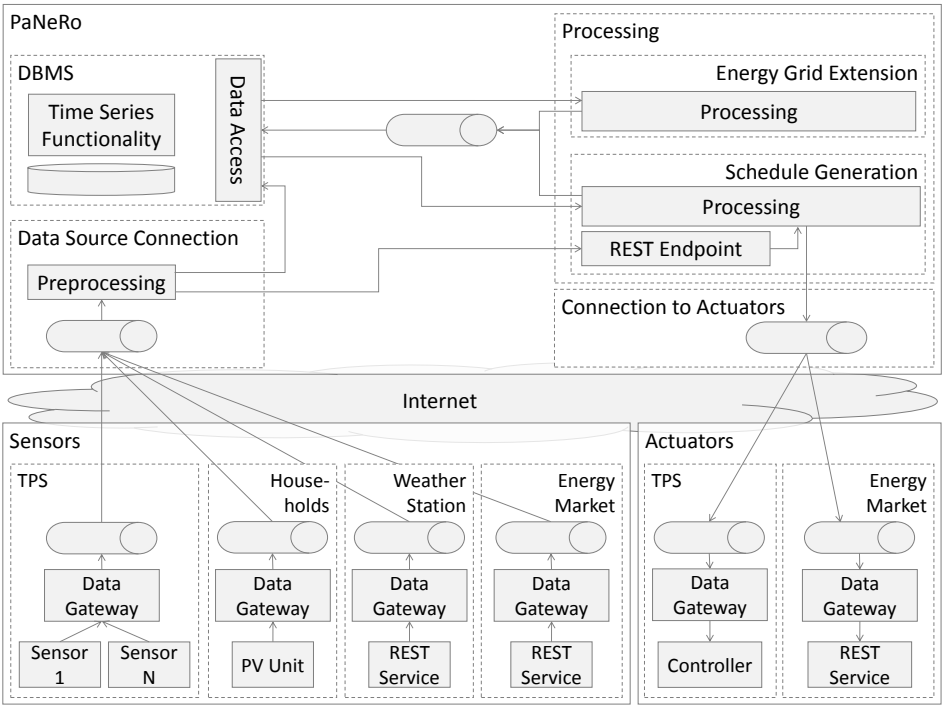


Fig. 3: Proposed Platform Architecture

security (R2a) and optionally also reliability (R2c), which implies that protocols such as the Message Queuing Telemetry Transport (MQTT; [OA14]) are not feasible in such cases.

Another option to accomplish security, loose coupling, and reliability, is to use Web Service technology with WS-Security and WS-ReliableMessaging [We05]. Typically, an enterprise service bus [Jo07] is used to implement loose coupling in the context of Web Services. Although Web Services offer a higher degree of interoperability especially because of the availability of more clients, we claim that message queuing systems offer enough interoperability in the context of the DMA, while keeping the overall architecture more lightweight and less complex.

A database system is chosen to satisfy R3a (scaling data store). The current requirements do not impose a certain type of data base system. Therefore, any type of database system (relational or NoSQL [Gr13]) can be chosen. For triggering an optimization (R3b), an analysis of the inbound data has to be implemented, which is typically addressed by a complex event processing (CEP) system [Lu02]. The current requirements do not impose selection criteria for a concrete CEP system.

The usage of a database system enables the optimization algorithms to receive the data in the required format (R4a), because format transformations can be implemented near the

database, while a selection criterion for a proper database system is to provide functions for time series calculations natively (R4b).

Although the requirements leave it open whether the optimization components should be placed near the database or can reside in a separate container, PANERO should offer a plugin possibility to directly deploy optimization algorithms (R5). This might be realized by stored procedures or other ways of tightly coupling implementations and the database system to benefit from a cheap data transfer between the database and the algorithm.

The connection to the actuators is done using a MQS, similar to the connection of sensors. When securing the connections, for example via transport layer security (TLS), this setting ensures security (R6a) and reliability (R6b). The concrete message format is not yet specified. When designing the concrete format, it will be able to cover all commands offered by the actuators and will be extensible to be able to express new commands (R6c). To guarantee that commands are available at actuators within specified timeframes the chosen MQS has to offer real-time capabilities (R6d). This might be the ActiveMQ Real Time sub-project [Ap15].

The data gateway used at the actuators enables a plausibility check of the commands (R7). This functionality has to be implemented for each actuator regarding to their operational constraints.

6 Cloud-readiness of the Platform

Cloud computing is a new innovative paradigm for offering and using IT resources: Resources can be used on demand and users do not need to care about the underlying infrastructure [Le09]. The primary reason for using Cloud technology is the possibility to move from capital to operational expenditure, which promises to save money [Ar10]. Currently, there are four different Cloud offerings available which can be distinguished as follows: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), and Composite as a Service (CaaS) [Le09, MG11]. In the case of IaaS, hardware is virtualized and the user has to install the software components by himself. On the contrary PaaS, software components such as databases or message queue systems are offered by the Cloud provider and the user does not need to take care about maintenance or scalability. However, if a database is offered in the Cloud, some authors claim that this is not a PaaS offering, but a separate service called Database as a Service—DbaaS [SK11]. Nevertheless, we follow the categorization by den Haan [dH13] and consider a database offering as a PaaS offering. In the case of SaaS, a complete software solution is offered as service. If, however, services are combined aiming to provide a new service in the Cloud, this is called CaaS.

The architecture of PANERO allows for deploying its components in the Cloud and, thus, makes use of the benefits of Cloud computing. The Data Source Connection is provided as a virtual machine image, which can be deployed multiple times depending on the number of incoming messages per second. For the database, a database PaaS offering can be used. Here, the scaling, backup and recovery are ensured by the Cloud provider. The processing

itself can be hosted on a virtual machine. The whole system will be offered as Composite as a Service to enable distribution grid providers to setup the platform when required without the need of manual intervention. For the automatic setup, the Topology and Orchestration Specification for Cloud Applications (TOSCA) standard is well suited [Bi12] as it provides a standardized way to package applications and their management plans into so called cloud service archives (CSAR). TOSCA provides a means to describe the components of a Cloud application, i. e., its composite structure, in a Cloud provide agnostic way. Summarized, TOSCA enables to easily deploy applications to different Cloud environments.

7 Related Work

For big data analysis, the Lambda Architecture [MW15] is a common pattern distinguishing two layers for the analysis: First, the batch layer to compute results based on the complete set of data. Second, the speed layer to provide results based on new data not yet available for the batch layer. The Lambda Architecture is applied in the approach presented by Strochbach et al. [St15] performing power quality analytics, real-time grid monitoring, and forecasting energy demand. Real world application is shown in the projects Big Data Public Private Forum (BIG) [BI15] and Peer Energy Cloud [PE15]. For the speed layer, they use a CEP system [Lu02]. In our architecture, there is no need for a speed layer as the processing algorithms always query the most recent data from the database. A CEP system may be used at the data source connection, where exceptional events can trigger an analysis at the processing. In the Peer Energy Cloud project, users can access their metered values. Security in that context is discussed by Ebinger et al. [Eb13]. PANERO currently does not yet offer its users a dashboard, but might include one and then will use the concepts presented in that paper.

Busemann et al. [Bu13] present a platform for processing readings of individual devices deployed at households. Although being similar to our architecture, the main difference is our employment of loose coupling between the sensors and the platform, and focusing on data originating from metering devices of a whole distribution grid.

In accordance to our proposal, Hauer et al. [Ha13] also see challenges regarding the shift towards renewable energies. They present an architecture for a future energy distribution system, but differ again in the use of loose coupling technologies.

Gustafsson et al. [Gu14] use the principles of Service-oriented Architecture (SOA) in district heating substations. They use SOA principles and the Web Service technology to connect sensors with a server. Our approach, in contrast, does not offer services using Web Service technology, but establishes loose coupling using message queuing technology.

IEC TR 62357 [IE12] provides an overview on the standards by the IEC technical committee 57 (“Power systems management and associated information exchange”). In our architecture, these standards may be used for the communication of the sensors with the data gateway where required. For the communication with PANERO, we use a MQS and a generic message format to reduce the complexity of heterogeneous data formats of the

sensors. Since PANERO only has to handle one well defined data format, this method saves additional effort.

8 Conclusion and Outlook

Motivated by the idea of a Decentralized Marked Agent, we derived requirements on an IT platform supporting the DMA. Based on the requirements, we outlined an architecture for the platform. The current requirements, however, do not explicitly point towards a certain type of a database system. As a consequence, we are recently in the process to evaluate whether relational or NoSQL databases are more applicable for the implementation of the data store within the platform. We did not discuss issues on connecting energy markets. A concrete description of the implementation of the respective data gateway will be provided as future work. The next step is to implement the platform and evaluate it in real-world settings of two different distribution grids. Our current research state indicates that the proposed short term optimization can be implemented by algorithms using the capabilities of a time series data base. Other research results indicate that it is also possible to use complex event processing [JZ14]. Thus, we will compare our implemented solution with CEP-based solutions. We mentioned the requirements on the security of smart meters [Fe14a], but did not do an evaluation of the proposed realization. This will be part of our future work after a prototype has been implemented.

Acknowledgments. The research leading to these results has received funding from the German government through the BMWi project NEMAR (03ET4018).

References

[Ap15] Apache Software Foundation: ActiveMQ Real Time, 2015. `http://activemq.apache.org/activemq-real-time.html`.

[Ar10] Armbrust, Michael et al.: A View of Cloud Computing. Commun. ACM, 53(4):50–58, April 2010.

[Bi12] Binz, Tobias; Breiter, Gerd; Leymann, Frank; Spatzier, Thomas: Portable Cloud Services Using TOSCA. IEEE Internet Computing, 16(03):80–85, May 2012.

[BI15] BIG: BIG Project Website, 2015. `http://www.big-project.eu/`.

[BM10] BMWi: Energiekonzept für eine umweltschonende, zuverlässige und bezahlbare Energieversorgung, 2010.

[Bu13] Busemann, C. et al.: Integrating Sensor Networks for Energy Monitoring with Service-Oriented Architectures. International Journal of Distributed Sensor Networks, 2013:1–12, 2013.

[dH13] den Haan, Johan: The cloud landscape described, categorized, and compared, October 2013. `http://www.theenterprisearchitect.eu/blog/2013/10/12/the-cloud-landscape-described-categorized-and-compared`.

[Eb13] Ebinger, Peter et al.: Privacy in Smart Metering Ecosystems. In: Smart Grid Security. Springer, 2013.

[Fe14a] Federal Office for Information Security: Protection Profile for the Security Module of a Smart Meter Gateway (Security Module PP), December 2014.

[Fe14b] Fehling, Christoph et al.: Cloud Computing Patterns: Fundamentals to Design, Build, and Manage Cloud Applications. Springer, 2014.

[Gr13] Grolinger, Katarina et al.: Data management in cloud environments: NoSQL and NewSQL data stores. J Cloud Comput Adv Syst Appl, 2(1):22, 2013.

[Gu14] Gustafsson, Jonas et al.: Application of Service Oriented Architecture for Sensors and Actuators in District Heating Substations. Sensors, 14(8):15553–15572, Aug 2014.

[Ha13] Hauer, I. et al.: Concept, architecture and components of a smart distribution control center. In: PowerTech. IEEE, 2013.

[HW03] Hohpe, G.; Woolf, B.: Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions. Addison-Wesley Professional, 2003.

[IE02] IEC: IEC 62056-21:2002 – Electricity metering – Data exchange for meter reading, tariff and load control - Part 21: Direct local data exchange, 2002.

[IE12] IEC: IEC TR 62357-1:2012 Power systems management and associated information exchange - Part 1: Reference architecture, 2012.

[IE13] IEC: IEC TR 61850-1:2013 Communication networks and systems for power utility automation - Part 1: Introduction and overview, 2013.

[Jo07] Josuttis, Nicolai M.: SOA in Practice: The Art of Distributed System Design. 2007.

[JZ14] Jerzak, Zbigniew; Ziekow, Holger: The DEBS 2014 grand challenge. In: DEBS. ACM Press, 2014.

[Le09] Leymann, Frank: Cloud Computing: The Next Revolution in IT. In: Proc. 52th Photogrammetric Week. Wichmann Verlag, pp. 3–12, September 2009.

[Lu02] Luckham, David: The Power of Events. Addison-Wesley, 2002.

[MG11] Mell, Peter; Grance, Timothy: The NIST Definition of Cloud Computing. online, October 2011. SP 800-145.

[MW15] Marz, Nathan; Warren, James: Big Data: Principles and best practices of scalable realtime data systems. Manning Publications, 2015.

[OA14] OASIS: MQTT Version 3.1.1, OASIS Standard, 2014.

[Op15] OpenMUC: OpenMUC, 2015. `http://www.openmuc.org`.

[Or00] Oracle Corporation: Function Shipping: Separating Logical and Physical Tiers, 2000. `http://docs.oracle.com/cd/A87860_01/doc/appdev.817/a86030/adx16nt4.htm`.

[PE15] PEC: Peer Energy Cloud project website, 2015. `http://www.peerenergycloud.de/`.

[SK11] Seibold, Michael; Kemper, Alfons: Database as a Service. Datenbank-Spektrum, 12(1):59–62, November 2011.

[St15] Strohbach, Martin et al.: Towards a Big Data Analytics Framework for IoT and Smart City Applications. Modeling and Processing for Next-Generation Big-Data Technologies, pp. 257–282, 2015.

[Th15] Thomson et al.: Darstellung des Konzepts – DMA Decentralised Market Agent – zur Bewältigung zukünftiger Herausforderungen in Verteilnetzen. In: Informatik. 2015.

[VHH12] Vossen, Gottfried; Haselmann, Till; Hoeren, Thomas: Cloud Computing für Unternehmen. dpunkt.verlag, 2012.

[We05] Weerawarana, Sanjiva et al.: Web Services Platform Architecture : SOAP, WSDL, WS-Policy, WS-Addressing, WS-BPEL, WS-Reliable Messaging, and More. Prentice Hall PTR, 2005.

All links were last followed on April 15, 2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

System Alarming für Energiesysteme

Michael J.M. Wagner[1], Friedrich Eisenhauer[2] und Dagmar Koß[3]

Abstract: System Alarming ist eine systemunterstützende Funktion, die bereits seit Jahrzehnten in den Telekommunikationsnetzen für störungsfreien Betrieb sorgt. Diese Funktion sammelt Informationen über abnorme Betriebszustände und ermöglicht es so dem Wartungspersonal zielgerichtet einzugreifen, bevor es zu einem Systemversagen kommt. Durch das Wichtigerwerden von erneuerbaren Energien werden unsere Energiesysteme immer heterogener. Eine Alarming-Implementierung mit offenen Schnittstellen ist daher eine Voraussetzung zur Abbildung eines Systemzustands. Angelehnt an die in der Telekommunikationsbranche genutzten Standards wird hier eine leichtgewichtigere Implementierung vorgestellt, die für den Einsatz in Energiesystemen geeignet ist. Exemplarisch wird diese Anwendung am fortiss-Demonstrator eingesetzt.

Keywords: Systemüberwachung, Entstörung, Alarmierung, Smart Grid

1 Einleitung

Die Zuverlässigkeit von komplexen Systemen beruht auf zwei Säulen: Zum einen müssen die Systeme selbst eine hohe Qualität aufweisen, zum anderen müssen Störungszustände einzelner Komponenten erkannt und behoben werden, bevor es zu einem Systemversagen kommt. System Alarming, kurz „Alarming“, ist eine systemunterstützende Funktion, die in vielen Bereichen verteilter Systeme zum Einsatz kommt. Aufgabe ist es Fehlerzustände, die im System erkannt werden, geeignet zu speichern, weiterzugeben und zu bilanzieren. Alarming wurde bereits vor etwa 20 Jahren in der Telekommunikationsbranche über Standards definiert und in verschiedenen Systemen implementiert.

Die Realisierung von Alarming im Telekommunikationsumfeld ist durch die Q3-Schnittstelle, definiert im ITU-Standard (International Telecommunication Union) [X.720] gekennzeichnet. Die Realisierung der Q3-Schnittstelle ist durch das Kommunikationsmuster Manager-Agent bestimmt. Jeder Knoten ist Agent für eine übergeordnete Management-Funktion, kann aber auch in komplexeren Zusammenhängen bezüglich untergeordneter Einheiten selbst als Manager auftreten [X.701].

Dieser Beitrag stellt die Alarmingkonzepte der Telekommunikation vor und zeigt, wie diese, entsprechend angepasst, für den Bereich der Energiesysteme ihre Nützlichkeit entfalten.

[1] WagnerTech UG, Turfstr. 18a, 81929 München, michael@wagnertech.de
[2] WagnerTech UG, Turfstr. 18a, 81929 München, fridoe@wagnertech.de
[3] fortiss GmbH, Guerickestraße 25, 80805 München, koss@fortiss.org

2 System Alarming in der Telekommunikationsbranche

2.1 Managed Object

Im ITU-Standard [X.700] wird *Managed Object* als generische Entität beschrieben, auf die alle administrativen Vorgänge bezogen sind. Die betriebsrelevanten Systemresourcen (HW/SW) werden in *Managed Objects* abgebildet. Zu Alarmierungszwecken haben diese optional einen Alarmstatus und können durch Notifikationen weitere Alarmparameter anzeigen, die im Standard [X.733] zum *Alarm Reporting* beschrieben werden.

Gegenstand der beiden folgenden Standards [X.734] und [X.735] ist, wie mit den Notifikationen bei auftretenden Ereignissen zu verfahren ist. In X.734 wird das *Logging* von Ereignissen (*Notification → Log Record*), in X.735 die Weiterleitung von Ereignissen (*Notification → Event Report*) beschrieben. Die Fähigkeit zur aktiven Weiterleitung von Ereignismeldungen (*Event Reports*) ist ein wesentlicher Bestandteil der Agent-Rolle im Q3-Protokoll. Gesteuert wird diese Weiterleitung durch den *Event Forwarding Discriminator*.

2.2 Alarm Summary

Die in den Standards X.733 bis X.735 beschriebenen Funktionalitäten beschreiben die Verwaltung von Alarmzuständen am einzelnen Objekt und deren *Logging* und Weiterleitung. Bei großen Systemen ist es aber nötig, aus Hunderten von Alarmzuständen eine einheitliche *Alarm Summary* (Alarmbilanz) zu erstellen. Diese Funktionalität ist in [Q.821] beschrieben. Alarmbilanzen lassen sich bezüglich verschiedener Gesichtspunkte erstellen, z.B. nach Funktionsgruppen oder den Hardwareeigenschaften der überwachten Objekte. Kernpunkt ist die Verwaltung der Dringlichkeiten, die es dem Entstörpersonal ermöglicht, schnell die richtigen Maßnahmen zu ergreifen. Zur in Q.821 beschriebenen Funktionalität gehört auch die Erstellung von periodischen Summary Reports.

3 Einsatz in Energiesystemen

Moderne Energiesysteme, in denen viele heterogene und dezentrale Einheiten zusammenwirken, stellen uns vor ähnliche Probleme, wie die Telekommunikationsnetze: Fehlfunktionen einzelner Komponenten werden von der Netzregelung kompensiert. Fehlt aber ein Alarming, das für eine gezielte Entstörung genutzt wird, kann es beim Ausfall zu vieler Komponenten zum Netzversagen kommen, was bei Energienetzen drastischere Auswirkungen hat, als in der Telekommunikation.

Die in den ITU-Standards beschriebenen Funktionalitäten sind teilweise etwas schwergewichtig. Die Q3-Objekte weisen eine hohe Komplexität auf. So enthalten beispielsweise der EventForwardingDiscriminator und das Log im Standard einen rekursiven Filter auf

alle im Managemtmodell befindlichen Objekte. In hochverteilten Systemen mit oftmals sehr kleinen Einheiten bietet sich daher ein leichtgewichtigerer Ansatz an. Das SNMP (*simple network management protocol*) der TCP/IP-Suite bietet Möglichkeiten, die in den Standards X.733 bis X.735 beschriebenen Funktionalitäten zu substituieren. Andere Protokolle bieten ähnliche Möglichkeiten Alarme zu melden. Die zentrale Funktion der Alarmbilanzierung ist mit diesen Protokollen aber nicht abgedeckt.

3.1 Systemaufbau

Die hier vorgestellte Software orientiert sich bei der Alarmbilanzierung am ITU-Standard Q.821, ohne alle Details zu implementieren. Das System hat folgenden Aufbau:

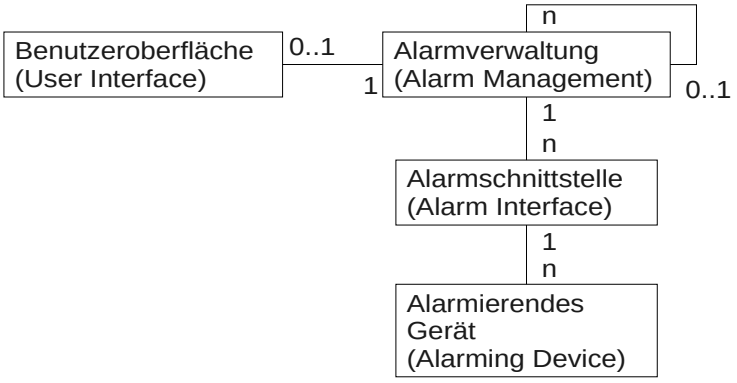


Abbildung 1: Systemaufbau

Die zentrale Einheit von Alarming ist die Alarmverwaltung. Diese kann hierarchisch aufgebaut und auf verschiedene Knoten verteilt sein. An einer Alarmverwaltung kann, muss aber nicht, eine Benutzeroberfläche angebunden sein. Die alarmierenden Geräte sind über Alarmschnittstellen, die ggf. eine Protokollumsetzung vornehmen, an die Alarmverwaltung angeschlossen.

Die Aufgaben der Alarmverwaltung sind:

- Aufzeichnung der Alarme: Jedes Alarmereignis, das über die Alarmschnittstelle gemeldet wird, wird in einer Logdatei gespeichert. Diese Funktion spiegelt das Logging aus X.734 wieder.
- Bewertung der Alarme: In Abhängigkeit von einer im alarmierenden Gerät selbst erkannten Beeinträchtigung der Gesamtfunktionalität (*service influence*) und dem Alarmtyp bestimmt eine Bewertungslogik auf der Basis von hinterlegten Regeln die Alarmpriorität.

- Alarmweiterleitung: Da Instanzen der Alarmverwaltung hierarchisch angeordnet sein können, entscheidet die Alarmweiterleitung über ein hinterlegtes Regelwerk, welche weitere Instanzen der Alarmverwaltung zu informieren sind. Die Weiterleitung kann auch an andere Ausgabemedien, wie e-mail, erfolgen.
- Bilanzierung: Um über eine große Anzahl von Alarmen einen Überblick zu erhalten, ist es wichtig, diese zu bilanzieren. Ein Bilanzmonitor überwacht bestimmte Alarmtypen und zeigt die anliegenden Alarmprioritäten. Falls eine höhere Alarmpriorität neu hinzukommt, stellt dies selbst wieder ein Alarmereignis dar, das an eine hierarchisch höher liegende Alarmverwaltung weitergeleitet werden kann.
- Alarmberichte: Der Bilanzmonitor ist das Einstiegsobjekt zur Entstörung: Ausgehend von einer Monitorinstanz kann sich das Wartungspersonal die Einzelalarme auflisten lassen.
- Datenpflege: Über die Datenpflege können die Regelwerke für die Bewertung und Weiterleitung von Alarmen gepflegt werden. Darüber hinaus können Bilanzmonitore mit den zu überwachenden Alarmtypen angelegt oder gelöscht werden.

3.2 Datenmodell

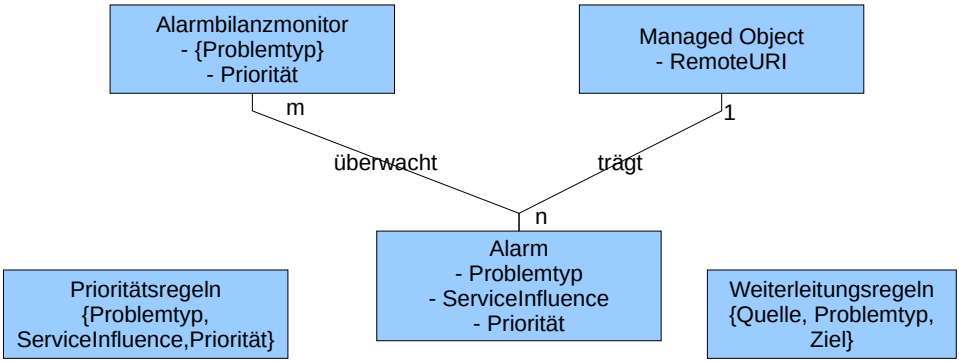


Abbildung 2: Datenmodell von System Alarming

Jeder Systemalarm wird genau einem Managed Object zugeordnet, dem Träger des Alarms. Ein Bilanzmonitor überwacht die anliegenden Alarme anhand einer Auswahlregel für Problemtypen. Er ermittelt die höchste anliegende Alarmpriorität. Die Prioritätsregeln ermöglichen die Ermittlung einer Alarmpriorität anhand der Parameter *Problemtyp* und *Service Influence*. Die ermittelte Alarmpriorität wird im Alarm hinterlegt. Die

Weiterleitungsregeln bestimmen, an welche weiteren Instanzen der Alarmverwaltung oder an welche e-mail-Adressen ein Alarmereignis weitergeleitet werden soll.

Statt der *Service Influence* kann auch ein Schwellwertpegel hinterlegt werden. In diesem Fall kann das System bei Schwellwertmeldungen entscheiden, welche Alarmstufe erreicht ist. Beim Unterschreiten des entsprechenden Niveaus wird wieder deeskaliert.

3.3 Einbettung in den fortiss-Demonstrator

Die Integration der Erzeugung von erneuerbaren Energien in die Haustechnik verwandelt letztere in Smart Grids. Damit die Haustechnik aber zu einem stabilen Regelsystem wird, das Produktion, Speicherung und Verbrauch der Energie steuert, wird eine skalierbare Software benötigt, die für verschiedene Größenordnungen von Systemen geeignet ist. Für den fortiss-Demonstrator[4] wurde eine mehrschichtige, SOA (*service oriented architecture*)-basierte und verteilte Softwarearchitektur gewählt, deren Implementierung offene Schnittstellen für anzuschließende Geräte auf der einen und Verwaltungssoftware auf der anderen Seite bietet. Verbunden werden Geräte- und Verwaltungsschicht durch eine Kommunikationsschicht. System Alarming wurde als ein Element der Verwaltungsschicht an die Kommunikationsschicht über eine WebService-Schnittstelle angebunden [Ko2012] (Abb. 3).

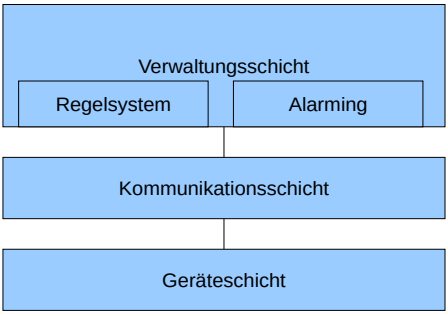


Abbildung 3: Architektur des fortiss-Demonstrators

Die Verwaltungsschicht enthält neben dem Alarming auch ein Regelsystem. Wie in [Du2014] beschrieben dient dieses Regelwerk zur Abstimmung der Energieerzeuger und -verbraucher. Hier können entsprechende Ereignisse auch der Alarmierungskomponente zugestellt werden.

Abbildung 4 zeigt den für die Pilotanwendung am fortiss Demonstrator realisierten Nachrichtenfluss.

[4] Projekt-Homepage: http://ses.fortiss.org/

Ein alarmierendes Gerät liefert seine Zustandsdaten über ein geräteabhängiges Protokoll an das Regelsystem. Über dieses Regelwerk können abnorme Betriebszustände erkannt und an System Alarming gemeldet werden.

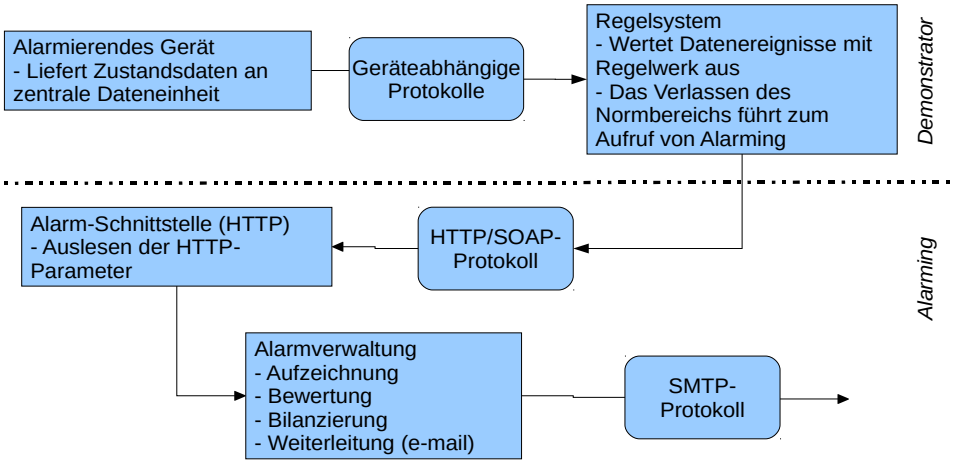


Abbildung 4: Nachrichtenfluss am fortiss Demonstrator

Die Alarmverwaltung zeichnet das Ereignis auf und bewertet den Alarm. Ist ein passender Bilanzmonitor eingerichtet, so ändert sich gegebenenfalls dessen Alarmbilanz. Erhöht sich dadurch die anliegende Alarmpriorität im Bilanzmonitor, so stellt dies selbst ein Alarmereignis dar. Über die Weiterleitungsregeln wird entschieden, ob das primäre Ereignis und/oder das Bilanzereignis weitergeleitet wird. Ziel der Weiterleitung kann eine weitere Instanz der Alarmverwaltung oder auch eine einfache e-mail-Adresse sein.

Im derzeitigen Integrationsstand ist die Eignung dieser System Alarming Imlementierung verteilte Ressourcen über eine hierarchische Struktur zu überwachen noch nicht genutzt. Dies soll in einem zweiten Schritt durch Anbindung eines weiteren Smart Grids erfolgen.

3.4 Benutzerführung

Fehlererkennung, -weiterleitung, -bilanzierung und -anzeige dienen schließlich dem Zweck, das Wartungspersonal bedarfsgerecht zum alarmierenden Gerät zu führen. Daher zeigt die Benutzeroberfläche als erstes eine Übersicht über die Bilanzmonitore mit ihren Alarmprioritätsbilanzen, die die Existenz von unbestätigten oder bestätigten Alarmen der jeweiligen Prioritäten und Alarmtypen aufzeigen (Abb. 5).

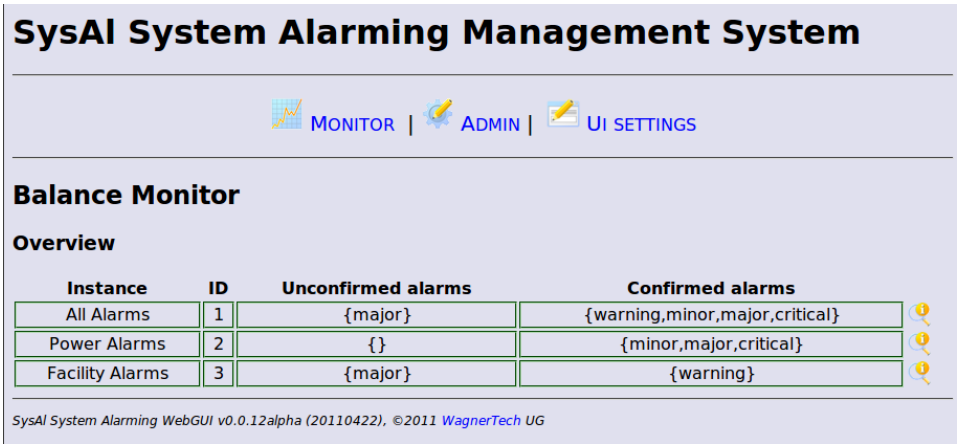


Abbildung 5: Einstiegsseite für die Alarmentstörung

Je nach Zuständigkeit des jeweiligen Wartungspersonals kann dieses dann die Alarme, die durch den betreffenden Monitor abgebildet werden, sich aufschlüsseln lassen und entstören. Die Unterscheidung von unbestätigten und bestätigten Alarmen ermöglicht es darüber hinaus, das Neueintreffen eines Alarms seit der letzten Alarmbestätigung zu erkennen.

4 Fazit

Die hier vorgestellte Implementierung ermöglicht in großen verteilten Systemen eine generische Alarmverwaltung. Die kennzeichnenden Eigenschaften sind die Hierarchisierbarkeit, die Möglichkeit der Alarmzusammenfassung über Bilanzmonitore, sowie die Ermittlung von Alarmprioritäten und Weiterleitungszielen über administrierbare Regelwerke.

In einer Pilotanwendung am fortiss Demonstrator werden verschiedene Elemente eines Energiesystems, wie Photovoltaikanlage, Batteriepuffer, sowie eine Klimaanlage als Verbraucher über ein Regelwerk verbunden. Über das Regelwerk können Systemzustände außerhalb des Normbereichs erkannt und über eine Serviceschnittstelle an System Alarming gemeldet werden. Die Verwaltung der Alarme, sowie der Einstieg zur Entstörung erfolgt über die in dieser Arbeit beschriebene Implementierung und dient der zielgenauen Entstörung zur Aufrechterhaltung der Gesamtfunktionalität.

Literaturverzeichnis

[Du2014] Duchon, M. et al.: Advancement of a Sensor Aided Smart Grid Node Architecture. CyperC Conference 2014.

[Ko2012] Koß, D. et al.: Establishing a Smart Grid Node Architecture and Demonstrator in an Office Environment Using the SOA Approach. SE4SG Workshop @ ICSE 2012.

[Q.821] ITU-T Recommendation Q.821: Stage 2 and stage 3 description for the Q3 interface – Alarm surveillance, 2000.

[X.700] ITU-T Recommendation X.700: Management Framework for Open Systems Interconnection (OSI) for CCITT Applications, 1992.

[X.701] ITU-T Recommendation X.701: Information technology - Open Systems Interconnection - Systems management overview, 1997.

[X.720] ITU-T Recommendation X.720: Information technology - Open Systems Interconnection - Structure of management information: Management information model, 1992.

[X.733] ITU-T Recommendation X.733: Information technology - Open Systems Interconnection - Systems Management: Alarm reporting function, 1992.

[X.734] ITU-T Recommendation X.734: Information technology - Open Systems Interconnection - Systems Management: Event report management function, 1992.

[X.735] ITU-T Recommendation X.735Information technology - Open Systems Interconnection - Systems Management: Log control function, 1992.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

IT und Dienstleistungen für die Energiewende und die Elektromobilität (IDEE 2015)

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

IT und Dienstleistungen für die Energiewende und die Elektromobilität (IDEE 2015)

Martin Matzner[1], Daniel Beverungen[1], Sebastian Lehnhoff[2], Gerhard Satzger[3]
und Michael Sonnenschein[4]

Die Energiewende mit ihren Zielen für die Umstellung des Energiesystems auf erneuerbare Energieträger bis zum Jahr 2050 führt zu einschneidenden Veränderungen in bisherigen Energieversorgungs- und Mobilitätssystemen. War die Energieerzeugung bisher im Wesentlichen am Bedarf ausgerichtet, so wird man in Zukunft vermehrt versuchen, alle möglichen Flexibilitäten im Verteilernetz zu heben und den Bedarf an das volatile Dargebot der erneuerbaren Energien anzupassen. Die Elektromobilität bietet dabei gute Voraussetzungen durch die Nutzung von Elektroautos als flexible Speichermedien. Die mit der Energiewende verbundenen politischen Vorgaben erfordern von bestehenden Akteuren im Energie- und Mobilitätsbereich mehr denn je, sich innovativen, dienstleistungsspezifischen Geschäftsmodellen zu öffnen. Bei diesem Transformationsprozess spielen daher nicht nur technologische Aspekte, sondern auch die Einbindung der Endverbraucher, die zunehmend die Rolle von Prosumern einnehmen, eine wichtige Rolle.

Daher liegt das Augenmerk dieses Workshops nicht nur auf der Systemintegration von Kleinerzeugern, Speichern und flexiblen Verbrauchsanlagen, wie zum Beispiel beim Einsatz von Wärmepumpen oder Elektroautos als Energiespeicher, sondern auch auf der Wechselwirkung dieser mit den Nutzern dieser Komponenten. Die Energie- und Wirtschaftsinformatik steht hier nicht nur in der Verantwortung, die „Systemintelligenz“ zu liefern, sondern steht auch vor der Aufgabe, Methoden für die nachgelagerte „Systemkompetenz“ zu schaffen und zu orchestrieren.

Im Rahmen des Workshops werden daher innovative Informationssysteme und IT-gestützte Dienstleistungen für die Energiewende und die Elektromobilität sowie Methoden zur Entwicklung dieser Systeme und Dienstleistungen diskutiert. Die Beiträge analysieren aktuelle Konzepte aus Wissenschaft und Praxis, entwickeln innovative Lösungen für die Energiewende oder erweitern Verfahren und Instrumente aus der (Wirtschafts-)Informatik- und Dienstleistungsforschung für das Handlungsfeld „Energiewende und Elektromobilität“.

[1] WWU Münster, ERCIS, Leonardo-Campus 3, 48149 Münster,
{martin.matzner | daniel.beverungen}@ercis.uni-muenster.de
[2] OFFIS, Institut für Informatik, Industriestraße 6, 26121 Oldenburg, sebastian.lehnhoff@offis.de
[3] KIT, Karlsruhe Service Research Institute, Englerstraße 11, 76131 Karlsruhe, gerhard.satzger@kit.edu
[4] Carl von Ossietzky Universität, Department für Informatik, Uhlhornsweg 84, 26111 Oldenburg,
michael.sonnenschein@uni-oldenburg.de

Alle Einreichungen zu diesem Workshop sind durch mindestens zwei Experten begutachtet worden. Wir danken den folgenden Mitgliedern des Programmkomitees für ihre Mitwirkung im Begutachtungsverfahren des Workshops „IDEE 2015“:

- Prof. Dr. Rainer Alt, Universität Leipzig
- Prof. Dr. Barbara Dinter, TU Chemnitz
- Dr. Patrick Jochem, Karlsruher Institut für Technologie
- Prof. Dr. Ralf Knackstedt, Universität Hildesheim
- Rudolf Leisen, Bundesministerium für Bildung und Forschung
- Prof. Dr. Jan Marco Leimeister, Universität Kassel, Universität St. Gallen
- Matthias Löchte, Stadtwerke Münster GmbH
- Prof. Dr. Kathrin Möslein, Universität Erlangen-Nürnberg
- Prof. Dr. Peter Palensky, TU Delft
- Prof. Dr. Jens Pöppelbuß, Universität Bremen
- Prof. Dr.-Ing. Christian Rehtanz, TU Dortmund
- Dr. Michael Schermann, TU München
- Prof. Dr. Hartmut Schmeck, Karlsruher Institut für Technologie
- Dominik Schulte, P3 energy
- Prof. Dr. Thorsten Staake, Universität Bamberg
- Dr. Jan Sudeikat, Hamburg Energie GmbH
- Prof. Dr. Oliver Thomas, Universität Osnabrück
- Verena Wiget, TÜV SÜD AG – E-Mobility

Münster, Oldenburg und Karlsruhe, im Juni 2015

Martin Matzner, Daniel Beverungen, Sebastian Lehnhoff, Gerhard Satzger und Michael Sonnenschein
(Workshop-Leitung)

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Demand Side Management in Residential Contexts – A Literature Review

Carola Gerwig[1], Dennis Behrens[1], Helmut Lessing[1], Ralf Knackstedt[1]

Abstract: The digitalization in the energy sector offers a great potential to implement Demand Side Management (DSM) in residential buildings and to exploit the flexibility on the side of the consumers. With the possibility to shift energy loads, it becomes easier to use renewable energies efficiently. The increasing number of home devices which can be controlled and the increasing number of electric cars are favorable conditions to implement DSM. This paper provides an overview of the state of the art approaches for DSM in residential contexts. Therefore, a structured literature review is carried out. Thereby, 598 articles are analyzed and categorized via a concept matrix. The results of the existing approaches are examined and it is pointed out in which areas further research is needed. This regards inter alia the need for a publicly available reference data set which would allow scientific benchmarking.

Keywords: Demand Side Management, Demand Response, Literatur Review, Load Shifting, Micro Grid, Smart Grid

1 Motivation

Demand Side Management (DSM) is the concept of shifting demand in order to achieve a balanced utilization in energy grids. This concept exists since the 1980s (cf. [Ge85]). Now it has become increasingly popular due to the integration of renewable energies in the energy systems. The energy production of these is not controllable but depends on external factors like the solar radiation or wind force. In contrast, the demand of many appliances like a dishwasher is flexible and can be shifted within a determined scope. The goal of DSM is to shift load from times in which the energy demand is high to times in which the demand is low or energy can purchased for a favorable price.

DSM becomes feasible due to the integration of information and communications technologies in energy grids. With an increasing number of smart meters[2] in residential buildings it is not restricted to industrial contexts anymore, but a new potential for load shifting in residential contexts emerges. This relative new aspect of DSM is supported by the increasing possibilities of connecting devices and services in the home environment. Another important driver are electric vehicles. The growing number of plug-in hybrid electric vehicles (PHEVs) and electric vehicles (EVs) arises a new potential for load shifting. It can be assumed, that users won't mind at which hour at night the car is charged as long it is ready for takeoff in the morning.

[1] University of Hildesheim, Institute for Information Systems, Üniversitätsplatz 1, 31141 Hildesheim, gerwig@uni-hildesheim.de/ dennis.behrens@uni-hildesheim.de

[2] In Germany the installation of smart meters in new buildings has been forced since 2010 by law, cf. 21b Abs. 3a EnWG

To apply DSM in residential buildings, different approaches have been developed. They differ in the physical framework in which DSM is embedded, in the components which are optimized and in the optimization algorithm itself. Not all of the approaches can be embedded into every framework and not all of the optimization methods can consider every component. To get an overview of the approaches, a structured literature review is required. Thereby, it should be pointed out which methods are used in which framework and which results can be achieved. In doing so, it can be identified, which the next steps are for bringing DSM in residential contexts forward. The guiding research questions are:

- What methods are used for DSM in residential contexts?
- Which framework is chosen for this?
- Which components are integrated?
- What are the comparable results?
- What datasets are used for the evaluation?

2 Research Design

To answer the raised research questions, a structuring framework for a review according to [WW02] is used. This includes a specification of a search string, the choice of suitable databases and an evaluation method. Thereby we follow the concept of [Br09] and construct a concept matrix in order to categorize the results.

2.1 Search String

Firstly, the term “Demand Side Management” is part of the search string. It is complemented by “Demand Response”, which is sometimes used as synonym. As the research focus is on DSM methods in residential contexts, the terms “residential” and ”optimization” are added. The search string is formulated in English as it is the established scientific language. Additionally, it is complemented by the German terms.

A first query on Google Scholar results in more than 6000 articles. Many of them deal with DSM in the fields of water and gas or with the trading of electricity. As these articles are not in in the scope of the research question, they are excluded by adding the following words to the search string:

- water - stock - gas - market - wasser - börse - markt.

In the final search string the single strings were combined with AND connectors:

“Demand Side Management” OR “Demand Response” OR “Demand - Side - Management” OR “Lastverschiebung”) AND (“Optimization” OR “Optimierung”) AND (“residential” OR “Wohn*”) - water - wasser - börse - stock - gas - markt - market

2.2 Databases

One database suitable for the scope of this literature review is the electronic Library AISeL[3], which is a central repository for scientific research in the field of information systems. Furthermore, most research papers are trackable by Google Scholar[4]. Thus, it is used as well. The search is restricted to the time range 2010 –2015 as smart meters for residential buildings have been introduced to the market about 2010.

2.3 Evaluation Method

The articles found are evaluated stepwise by the title, the abstract and finally by the whole article. In this process, all articles are sorted out, which cannot contribute to the research questions. As relevant identified articles are evaluated and categorized via concept matrix. As a result, the different approaches can be compared more easily.

3 Concept Matrix

To classify the articles found systematically, a concept matrix is developed (cf. [Br09]). Therefore, the dimensions are determined by the research questions. One dimension is the physical *framework* in which a proposed method is embedded. One dimension are the *optimization methods*. For the evaluation of the methods, it is important to know which components are optimized and which assumptions are used for the databasis. These questions are captured in two more dimensions: *components* and *data basis*. The dimensions are visualized in figure 1.

3.1 Framework

There are different frameworks to control and motivate load shifting in residential buildings. A rough division can be made between locally controlled DSM and centralized controlled DSM. Moreover there are hybrid forms, in which a hierarchical structure of the smart grid is supposed. Another broad division concerns the communication, which can be one-way or bidirectional. Optionally, DSM can be performed automatically and/or manually. To develop a stringent but clear classification, four categories are used to cover the possible frameworks concerning DSM for residential loads.

The first category is *centralized controlled DSM*, which means that the deferrable loads of several households and/or EVs are controlled by one central node. This category includes the central management of several households and/or charging stations up to whole cities. Furthermore, the central node can be part of a larger hierarchical system. There has to be a bidirectional communication between the deferrable load and the controlling node.

[3] http://aisel.aisnet.org/
[4] http://scholar.google.de/

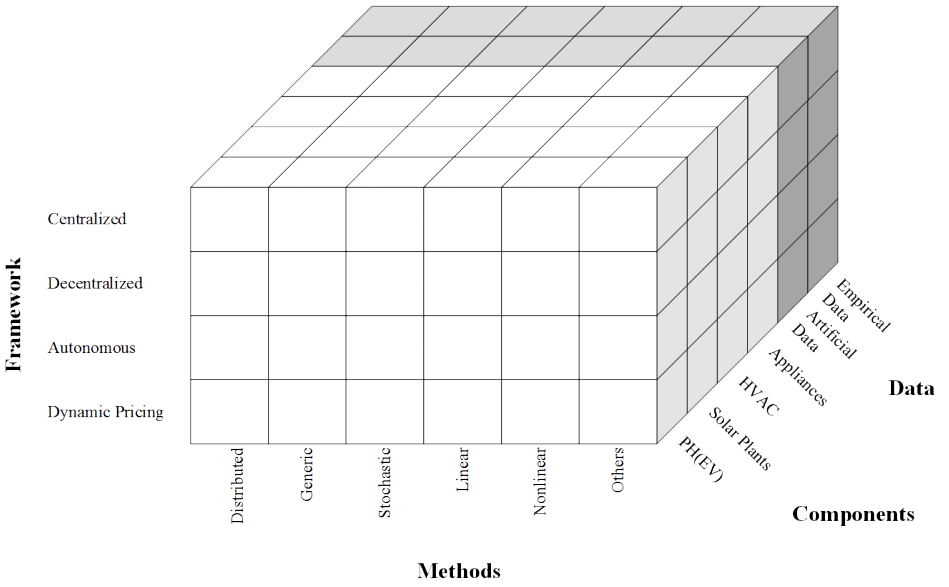


Fig. 1: Dimensions of the Concept Matrix

The second category is *distributed coordination of DSM*. In this scenario several households and/or charging stations communicate in a bi- or multidirectional manner to establish an optimal scheduling plan to match their demands.

Another approach is *autonomous DSM for one individual household*, possibly with one EV. The optimization problem only concerns the energy management of one household and one charging station. The goals are to increase the level of captive use, if a solar plant is installed, and/or to lower the energy bill by shifting loads to more favorable times. This approach can be combined with methods of the next category, which proposes variable energy prices as incentive for load shifting.

The last categoriy is *Dynamic Pricing*. The goal is not to control deferrable loads, but to motivate consumers to shift their load manually or to trigger smart devices by implementing variable electricity prices. In the process new peaks of energy demand have to avoided. The communication network is solely a one-way communication medium, through which the electricity prices are communicated.

3.2 Methods

In all cases a mathematical optimization problem has to be solved. There are many methods which can be applied. Considering the complex framework, heuristic methods in the fields of *genetic algorithms* and *stochastic optimization* can be expected, whereas *distributed algorithms* can support the distributed structure of the deferrable loads. Obviously,

standard solvers of the fields of mixed integer *linear* and *nonlinear programming* can also be expected.

3.3 Components

To get an overview, which methods can deal with which components, it is distinguished between four groups of components. The first one are *household appliances* which have a fixed load profile. Only the starting point is flexible. These appliances could be dish washers, washing machines, dryers etc. The second group describes *thermic loads* like heating and air conditioning (*HVAC*). They are characterized by the goal to reach or to hold a certain temperature. In doing so, energy is needed in more or less cyclic terms. The third group of consumer are *electric vehicles or plug–in electric vehicles*. In the charging process a certain amount of energy is needed. The electrical power can vary within certain bounds. Further components, which can optionally be considered, are *solar plants*. The energy production of these is not controllable, but it can be included in the constraints of the optimization problem.

3.4 Data Basis

In order to compare the results of the articles, it is important to have a look into the data sets used for evaluation. Thus it is reviewed, if they are *empirically obtained* or *artificially created*.

4 Results

4.1 Classification in the Concept Matrix

The keyword-based search leads to 598 articles[5]. A huge amount of articles could be sorted out by reading the titles and abstracts. These were for example articles about the management of storages, about DSM in commercial and industrial contexts or about the infrastructure without any optimization method. The remaining 55 articles[6] have been evaluated by reading the whole articles carefully. Thereby, 48 articles have been identified as relevant and have been categorized via the concept matrix (s. table 1). The evaluation is done stepwise for each framework. A comparison of the evaluations done in the papers is presented in the next subchapter. It is followed by the conclusions and limitations.

[5] The search was conducted between 23^{th} and 28^{th} December 2014

[6] For 5 articles, it has not been possible to organize the full version. These have been categorized by reading the abstract. They are marked with a *.

References	Framework				Methods						Components				Data	
	centralized	decentralized	autonomous	Dynamic Pr.	distributed	genetic	stochastic	linear	nonlinear	others	appliances	HVAC	solar plants	(PH)EV	empirical	artificial
[ASHK13]			X					X			X					X
[Al12]			X							X	X					X
[Al13]			X	X						X						X
[DA12]			X					X			X					X
[At13]			X							X	X					X
[AYK14]				X			X									
[Ba14]			X				X				X				X	
[BF13]	X								X					X	X	
[BLH12]		X			X									X		X
[BN14]			X							X	X				X	
[CC11]			X					X			X		X			X
[Ch13]		X			X									X		X
[Co14]	X							X							X	
[CYN14]		X			X					X	X			X		X
[Du13]		X						X						X	X	X
[Fe12]	X								X		X	X		X		X
[FG12]		X	X				X							X		X
[HMN12]		X							X		X					X
[Hu14]			X				X				X					X
[Ji11]			X				X				X	X		X		X
[KG11]			X							X	X					X
[Li14a]*			X			X					X					
[Kw12]		X					X					X			X	
[LF14]			X				X				X		X			X
[Li14b]			X				X				X					X
[LP13]			X			X					X					X
[LPK12]			X			X								X		X
[Ma14a]		X			X						X			X		X
[Ma12]			X							X						X
[Ma14b]*			X						X		X	X				X
[Me10]	X		X						X					X		X
[MH14]	X								X					X		X
[MHC12]			X			X					X					
[MM13]		X			X						X	X		X		X
[Mi11]			X					X				X	X			X
[Pe13]			X							X	X	X				X
[RZ15]*			X			X					X	X				
[Sa10]		X			X						X					X

References	Framework				Methods						Components				Data	
	centralized	decentralized	autonomous	Dynamic Pr.	distributed	genetic	stochastic	linear	nonlinear	others	appliances	HVAC	solar plants	(PH)EV	empirical	artificial
[SLG14]		X			X									X		X
[SXvdS14]	X				X			X								X
[St15]*			X							X						X
[Ta13]*	X					X								X		X
[DVF13]	X									X	X					X
[Ve14]		X					X							X		
[YTN12]				X	X											X
[Za13]	X						X				X					X
[ZBG12]*	X					X					X	X				
[Zh13]	X					X					X	X	X			
Summe	**11**	**12**	**25**	**3**	**9**	**8**	**9**	**8**	**6**	**10**	**26**	**11**	**4**	**16**	**6**	**36**

Tab. 1: Concept Matrx augemented with references

4.2 Evaluation of Concept Matrix

Autonomous DSM for one Individual Household

A large proportion (25 of 48) of the articles deals with autonomous DSM for individual households. Thereby, genetic algorithms, stochastic optimization and linear programming are each proposed by 20% of the articles. Further methods are pursuit algorithms ([Al12], [Ma12]), Multi-Objective-Programming [BN14], Fuzzy Clustering [Pe13], combinatorial mathematics [Al13] and Dynamic Programming [KG11].

Six articles consider thermic loads, which increase the complexity of the optimization problem. In [Ma14b] and [Mi11] these loads are modeled by non-linear constraints. [Ji11] uses a stochastic differential equation and optimizes the expected value. In [At13], a thermal energetic building simulation is used to generate training data for a neural network. It learns to keep the temperature. Three articles [CC11], [LF14], [Mi11] optimize the captive use of solar plants on the rooftop. Thereby, [LF14] takes the uncertainties in supply and generation into account by using confidence intervals.

Decentralized Organized DSM

Decentralized organized DSM is examined by 12 articles. Thereby, eight articles are built up on the model of Mohsenian-Rad et al. [MR10], in which distributed agents find the unique Nash-equilibrium by using best-response-algorithms. The equilibrium is equivalent to the global solution of the optimization problem, if certain conditions are met. In [Ma14a] and [SLG14] the model is extended by integrating energy storages. [Sa10] adapts the model for Real-Time-Pricing instead of determined cost functions. In this article and

in [Ch13] the iterations of the distributed algorithm are not performed sequentially, but simultaneously in order to speed up the calculation time. The approach of [CYN14] is based on game-theoretical considerations as well. To guarantee the convergence of the algorithm, a penalty term is used for changes of the scheduling plan. [BLH12] extends the model by adapting the constraints for EVs and by considering physical constraints of the energy grid. The problem of not knowing the constraining parameters of the optimization problem 24h in advance is addressed in [FG12]. Therefore, random variables are used. In [MM13] it is proven, that their proposed algorithm, which uses a division of the optimization problem, is three times faster than a centralized approach.

In [HMN12] priorities are assigned to the loads. The optimization problem is then solved by a distributed algorithm. [Ve14] models the optimization problem with Markov processes and thus the model can handle uncertainties. In [Du13] reinforcement learning is used to organize the charging process of EVs.

Thermic loads are only modeled by [Kw12] in contrary to electric vehicles, which are considered in the constraints by 9 of 12 articles.

Centralized Organized DSM

In the review process 11 articles are identified, which assign the management of residential loads to one central coordination point which coordinates more than one household. Four articles ([BF13], [Me10], [MH14], [Fe12]) apply nonlinear programming to solve the resulting complex optimization problem. Three articles favor genetic algorithms like [Zh13], [Ta13], [ZBG12]. In [Za13], there is a central agent, which calculates the load shifting by a fuzzy Markov game and simulated annealing. The actions of the consumers are captured by Q-Learning. In [SXvdS14] the optimization problem is modeled as game, in which a central node calculates the best actions for the other nodes, which represent appliances. [DVF13] optimizes HVAC systems with the scheduling algorithm Earliest Deadline First.

Half of the methods concentrate on (PH)EVs. To optimize the charging process [Zh13] uses a genetic algorithm, in [Ta13] the optimization uses an Estimation Distribution Algorithm. [BF13] puts up a quadratic optimization problem in which the size of the battery, the loading capacity and the current energy prices are considered. The evaluation is performed on the National Household Travel Survey (NHTS). [MH14] develops an online application which coordinates the charging process of electric vehicles.

Dynamic Pricing

Dynamic Pricing is an important aspect for DSM as many approaches react on price incentives. Variable energy tariffs are often part of the constraints for the optimization problem. Nevertheless, it is usually examined in larger context and does not focus on individual residential households. Furthermore, some papers may be not found by the keyword-search as the term ”market” is excluded. As a result there are only three articles about the design of dynamic energy prices found in the selection.

In [YTN12] the scheduling task is modelled as non-cooperative game. Assuming that consumers choose the action, which is most profitable for them, the energy prices are cal-

culated which lead to the most balanced load distribution. In [AYK14] the energy prices are determined one hour ahead by looking into the predicted loads and the probability of consumers reacting on price changes. [Al13] proposes a tariff which does not only change in time, but considers the maximal demand of the day in the billing process. In doing so, new load peaks as reaction to low prices can be avoided.

It is not specified, which deferrable loads are modeled in these approaches as it is the choice of the consumers. Nevertheless, concerning a simulation it would be important to model a realistic scenario.

4.3 Comparison of the Evaluations and Data Sets

Nearly all (41 of 48) articles provide simulation results to their approaches. About 70% of the articles generate artificial data sets. For example [SXvdS14] uses standard load profiles to simulate the load distribution for ten households. Thereby, 40% of the load is assumed to be flexible. In [Sa10] the demand of 250 households is modeled with a random distribution and it is assumed, that 50% of the loads are flexible. In [Ve14] the charging processes of 20 EVs at 20 households are modeled. Thereby, different percentages (15%, 30%, 45%) of flexible households are tested. The constraints are modelled by Gaussian distributions. In [LP13], [CC11] and [Ji11], the simulations only include deferrable loads. In [LP13], these loads are modeled with an inflexible load profile, which is simulated by a random distribution. In [CC11], a dish washer, a washing machine and a dryer are modelled in detail. Similar approaches have been used for the other articles with a artificially generated data set.

Six articles refer to empirically recorded data sets: In [Du13] the Irish Smart Metering Data from the Commission for Energy Regulation ([CE15] is used to simulate an optimal charging management of nine EVs connected to nine households. The evaluation of [Ba14] also derives the constraining parameter by using Smart Metering Data. It uses the Olympic Peninsula Demonstration Testbed [Ch10]. [BN14] applies the real energy prices which were charged in Long Island of New York 2013 [Ne13] to simulate the optimal usage of three household appliances (dishwasher, washing machine and dryer). A similar evaluation method is used in [Kw12]. Therefore, the total consumption of the University of Southern California as well as the consumption of single appliances like the HVAC, light and laptops were measured. Based on this, load shifting could then be simulated for a relatively realistic scenario. [BF13] looks into the charging management of EVs and uses the U.S. National Household Travel Survey Dataset [20] to simulate usage patterns. For [Co14] a field trial was carried out, in which 10 households took part in. The users specified, in which time frames they would like to do their laundry seven days ahead. A central node calculated the optimal starting points and informed the users via SMS.

The goal of the presented approaches is to reduce peaks in consumption or to adapt the consumption in accordance to a cost function. Thereby, a cost reduction of about 20 – 75% is achieved. However, the results are not comparable as different assumptions are made. This not only includes the cost functions but also the uncertainty of the forecasted loads,

the number of updates used for one day, the modeling of the constraining parameters and the percentage rate of flexible loads in the total consumption.

4.4 Conclusions

Particular emphasis of the research done so far is placed on autonomous DSM. The approach benefits from independence to other households. The optimization problem is limited to a manageable amount of appliances, which are partially modeled in detail. There are various optimization methods which can cope with this kind of problems. The communication infrastructure can be relatively simple as in most cases the only information which has to be communicated are the energy tariffs for the next 24 hours. However, the independence to other households has its drawbacks. Usually, the best solution for a local network load cannot be reached this way. Decentralized DSM pursuits this and allows negotiations of several households. By doing this, the number of degrees of freedom is increased and the parameters are connected to each other, so that a generation of new peaks can be avoided. Under certain conditions, a global optimum can be reached that way as [MR10] proves by applying game-theoretical theorems. Most of the research found for decentralized DSM relies on the model of [MR10] and tries to enhance the approach in certain aspects.

Given that all constraining parameters are known, a centrally organized DSM can find to the optimal solution. For example, in [Me10] a centralized strategy for charging EVs is compared to a local one, in which a charging process is controlled by the smart home management of one individual household. Thereby, the decentralized strategy reaches slightly better results, but it also requires a more complex communication infrastructure. The need of a multi-directional communication infrastructure is a drawback for all approach of DSM which include more than one household. Additionally, questions of data protection turn up as sensitive data is exchanged between the different participants. This might be the reason why 5 of 10 articles dealing with the optimization of EV charging concentrate on public stations. In this context, the question of data privacy is not as critical as in the private context.

In the evaluation of the methods proposed in the articles, crucial short-comings are revealed. The data sets used for the evaluations are often based on highly simplified assumptions and can therefore only be used for feasibility tests of the algorithms. To carry out scientific benchmarking publicly available data sets are needed which are measured empirically and in which deferrable loads are marked. Furthermore, it would be useful to know how long in advance the deferrable loads are reported.

Besides real data sets used for evaluation in [Du13] and [Ba14], data sets of Non-Intrusive-Appliance-Load-Monitoring (NIALM) can be recommended. There is for example the AMPds dataset ([Ma13]) in which the demand of each component is stated individually.

4.5 Limitations

The keyword search cannot guarantee that all articles related to DSM in residential context are found. To the contrary, some relevant articles are missing for sure. However, the present literature review gives a comprehensive overview of the methods and algorithms used in the residential context of DSM. In addition to that, short-comings and open questions are pointed out.

5 Summary and Outlook

The literature review clearly shows that the implementation of DSM in a residential context is a current research topic in which various different approaches are tested. By categorizing the articles for DSM in different dimensions, it can be seen which approaches work for which frameworks, for which components and with which optimization algorithms. The analysis of the evaluations done in the articles reveals a lack of empirically measured and publicly available data sets and an agreement on error measure to compare the performance of the algorithms.

Some articles take up the fact, that the input parameters of the assumed demand can only be random variables. It seems that further research is needed to cope with this fact. Additionally, it is not fully modelled yet, that deferrable loads are not known 24h ahead, but only reported during the day. Thus, there is a research gap concerning the explicit consideration of uncertainties and a continuously adaptation of the demand scheduling during the day.

References

[20] 2009 National Household Travel Survey, URL: http://nhts.ornl.gov Stand: 18.04.2015.

[Al12] Ali, S.Q. et al.: Pursuit Algorithm for optimized load scheduling. In: Power Engineering and Optimization Conference (PEDCO) Melaka, Malaysia, 2012 Ieee International. pp. 193–198, June 2012.

[Al13] Ali, SQ et al., NH: Load scheduling with maximum demand and time of use pricing for microgrids. In: Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS), 2013 IEEE. IEEE, 2013.

[ASHK13] Alam, M.R.; St-Hilaire, M.; Kunz, T.: Cost optimization via rescheduling in smart grids; A linear programming approach. In: Smart Energy Grid Engineering (SEGE), 2013 IEEE International Conference on. pp. 1–6, Aug 2013.

[At13] Atabay, Dennis et al.: Self-Adapting Building Models and Optimized HVAC Scheduling for Demand Side Management. 2013.

[AYK14] Ashfaq, Ahsan; Yingyun, Sun; Khan, Akif Zia: Optimization of economic dispatch problem integrated with stochastic demand side response. In: Intelligent Energy and Power Systems (IEPS), 2014 IEEE International Conference on. IEEE, pp. 116–121, 2014.

[Ba14] Bassamzadeh, Nastaran; Ghanem, Roger; Lu, Shuai; Kazemitabar, Seyed Jalal: Robust scheduling of smart appliances with uncertain electricity prices in a heterogeneous population. Energy and Buildings, 84:537–547, 2014.

[BF13] Bashash, Saeid; Fathy, Hosam K: Optimizing demand response of plug-in hybrid electric vehicles using quadratic programming. In: American Control Conference (ACC), 2013. IEEE, pp. 716–721, 2013.

[BLH12] Beaude, Olivier; Lasaulce, Samson; Hennebel, Martin: Charging games in networks of electrical vehicles. In: Network Games, Control and Optimization (NetGCooP), 2012 6th International Conference on. IEEE, pp. 96–103, 2012.

[BN14] Bu, Honggang; Nygard, Kendall: Adaptive Scheduling of Smart Home Appliances Using Fuzzy Goal Programming. In: ADAPTIVE 2014, The Sixth International Conference on Adaptive and Self-Adaptive Systems and Applications. pp. 129–135, 2014.

[Br09] vom Brocke, Jan; Simons, Alexander; Niehaves, Björn; Riemer, Kai; Plattfaut, Ralf; Cleven, Anne: Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: 17th European Conference on Information Systems, ECIS 2009, Verona, Italy, 2009. pp. 2206–2217, 2009.

[CC11] Castillo-Cagigal, Manuel et al.: A semi-distributed electric demand-side management system with PV generation for self-consumption enhancement. Energy Conversion and Management, 52(7):2659–2666, 2011.

[CE15] CER Smart Metering Project, URL: http://www.ucd.ie/issda/data-/commissionforenergyregulationcer/, Stand: 26.01.2015.

[Ch10] Chassin, D.P.: Olympic Peninsula Demonstration Testbed Results.Tech.Rep. 2010.

[Ch13] Chen, He Henry et al.: A variational inequality approach to instantaneous load pricing based demand side management for future smart grid. In: Communications (ICC), 2013 IEEE International Conference on. IEEE, pp. 4250–4254, 2013.

[Co14] Costanza, Enrico et al.: Doing the laundry with agents: a field trial of a future smart energy system in the home. In: Proceedings of the 32nd annual ACM conference on Human factors in computing systems. ACM, pp. 813–822, 2014.

[CYN14] Chavali, Phani; Yang, Peng; Nehorai, Arye: A distributed algorithm of appliance scheduling for home energy management system. Smart Grid, IEEE Transactions on, 5(1):282–290, 2014.

[DA12] De Angelis, Francesco et al.: Optimal task and energy scheduling in dynamic residential scenarios. In: Advances in Neural Networks–ISNN 2012. Springer, 2012.

[Du13] Dusparic, Ivana et al.: Multi-agent residential demand response based on load forecasting. In: Technologies for Sustainability (SusTech), 2013 1st IEEE Conference on. IEEE, pp. 90–96, 2013.

[DVF13] Della Vedova, Marco L; Facchinetti, Tullio: Real-Time Scheduling for Peak Load Reduction in a Large Set of HVAC Loads. In: ENERGY 2013, The Third International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies. pp. 161–166, 2013.

[Fe12] Fernandes, Filipe et al.: Intelligent Management of End Consumers Loads Including Electric Vehicles through a SCADA System. 2012.

[FG12] Fathi, Mohammad; Gholami, Mehrdad: Localized demand-side management in electric power systems. In: Smart Grids (ICSG), 2012 2nd Iranian Conference on. IEEE, pp. 1–4, 2012.

[Ge85] Gellings, Clark W: The concept of demand-side management for electric utilities. Proceedings of the IEEE, 73(10):1468–1470, 1985.

[HMN12] Huang, Yingsong; Mao, Shiwen; Nelms, RM: Smooth electric power scheduling in power distribution networks. In: Globecom Workshops (GC Wkshps), 2012 IEEE. IEEE, pp. 1469–1473, 2012.

[Hu14] Hu, Ting et al.: An optimal day-ahead dispatch strategy for deferrable loads. In: Power System Technology (POWERCON), 2014 International Conference on. IEEE, pp. 1785–1790, 2014.

[Ji11] Jia, Liyan et al.: Multi-scale stochastic optimization for home energy management. In: Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2011 4th IEEE International Workshop on. IEEE, pp. 113–116, 2011.

[KG11] Kim, Seung-Jun; Giannakis, Georgios B: Efficient and scalable demand response for the smart power grid. In: Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2011 4th IEEE International Workshop on. IEEE, pp. 109–112, 2011.

[Kw12] Kwak, Jun-young et al.: SAVES: A sustainable multiagent application to conserve building energy considering occupants. In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1. pp. 21–28, 2012.

[LF14] Liu, Ping; Fu, Yong: Construction of multiband uncertainty set for building energy management with uncertain loads and solar power. In: PES General Meeting— Conference & Exposition, 2014 IEEE. IEEE, pp. 1–5, 2014.

[Li14a] Liang, Ying et al.: Auto-DR Optimal Decision-Making Method of Smart Home Based on RTP. In: Applied Mechanics and Materials. volume 602. Trans Tech Publ, pp. 3817–3821, 2014.

[Li14b] Liu, Gi-Ren et al.: Optimal Threshold Policy for In-home Smart Grid with Renewable Generation Integration. 2014.

[LP13] Lee, Junghoon; Park, Gyung-Leen: Power load distribution for wireless sensor and actuator networks in smart grid buildings. international Journal of Distributed sensor Networks, 2013, 2013.

[LPK12] Lee, Junghoon; Park, Gyung-Leen; Kim, Hye-Jin: Multithreaded Power Consumption Scheduler Based on a Genetic Algorithm. In: Communication and Networking, pp. 47–52. Springer, 2012.

[Ma12] Maqbool, SD et al.: Comparison of pursuit and ε-Greedy algorithm for load scheduling under real time pricing. In: Power and Energy (PECon), 2012 IEEE International Conference on. IEEE, pp. 515–519, 2012.

[Ma13] Makonin, Stephen et al.: AMPds: A public dataset for load disaggregation and ecofeedback research. In: Electrical Power & Energy Conference (EPEC), 2013 IEEE. IEEE, pp. 1–6, 2013.

[Ma14a] Manasseh, Emmanuel et al.: Distributed demand-side management with load uncertainty. In: ITU Kaleidoscope Academic Conference: Living in a converged world-Impossible without standards?, Proceedings of the 2014. IEEE, pp. 23–28, 2014.

[Ma14b] Marco, Severini et al.: Smart home task and energy resource scheduling based on non-linear programming. In: Recent Advances of Neural Network Models and Applications, pp. 175–185. Springer, 2014.

[Me10] Mets, Kevin et al.: Optimizing smart energy control strategies for plug-in hybrid electric vehicle charging. In: Network operations and management symposium workshops (NOMS Wksps). pp. 293–299, 2010.

[MH14] Milano, Federico; Hersent, Olivier: Optimal load management with inclusion of electric vehicles and distributed energy resources. Smart Grid, IEEE Transactions on, 5(2):662–672, 2014.

[MHC12] Miao, Hui; Huang, Xiaodi; Chen, Guo: A genetic evolutionary task scheduling method for energy efficiency in smart homes. International Review of Electrical Engineering, 7(5):5897–5904, 2012.

[Mi11] Missaoui, Rim et al.: Energy Fluxes optimization for PV integrated Building. In: PowerTech, 2011 IEEE Trondheim. IEEE, pp. 1–7, 2011.

[MM13] McNamara, Paul; McLoone, Sean: Hierarchical Demand Response using Dantzig-Wolfe decomposition. In: Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES. IEEE, pp. 1–5, 2013.

[MR10] Mohsenian-Rad, A-H et al.: Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. Smart Grid, IEEE Transactions on, 1(3):320–331, 2010.

[Ne13] New York Independent System Operator: Pricing Data, URL: http://www.nyiso.com-/public/markets_operations/market_data/pricing_data/index.jsp.

[Pe13] Pereira, R et al.: Demand Response Analysis in Smart Grids Using Fuzzy Clustering Model. In: Technological Innovation for the Internet of Things, pp. 403–412. Springer, 2013.

[RZ15] Ranjini, A; Zoraida, BSE: Preemptive Appliances Scheduling in Smart Home Using Genetic Algorithm. In: Artificial Intelligence and Evolutionary Algorithms in Engineering Systems, pp. 387–393. Springer, 2015.

[Sa10] Samadi, Pedram et al.: Optimal real-time pricing algorithm based on utility maximization for smart grid. In: Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on. IEEE, pp. 415–420, 2010.

[SLG14] Soliman, Hazem M; Leon-Garcia, Alberto: Game-theoretic demand-side management with storage devices for the future smart grid. 2014.

[St15] Stüdli, Sonja et al.: Distributed Load Management Using Additive Increase Multiplicative Decrease Based Techniques. In: Plug In Electric Vehicles in Smart Grids, pp. 173–202. Springer, 2015.

[SXvdS14] Song, Linqi; Xiao, Yuanzhang; van der Schaar, Mihaela: Non-stationary demand side management method for smart grids. In: Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, pp. 7759–7763, 2014.

[Ta13] Tazoe, Takashi et al.: EDA-Based Charging Algorithm for Plug-In Hybrid Electric Vehicle to Shift the Peak of Power Supply. In: Applied Mechanics and Materials. volume 267. Trans Tech Publ, pp. 46–50, 2013.

[Ve14] Verschae, Rodrigo et al.: A distributed coordination framework for on-line scheduling and power demand balancing of households communities. In: Control Conference (ECC), 2014 European. IEEE, pp. 1655–1662, 2014.

[WW02] Webster, Jane; Watson, Richard T: Analyzing the past to prepare for the future: Writing a literature review. Management Information Systems Quarterly, 26(2):3, 2002.

[YTN12] Yang, Peng; Tang, Gongguo; Nehorai, Arye: Optimal time-of-use electricity pricing using game theory. In: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, pp. 3081–3084, 2012.

[Za13] Zang, Chuanzhi et al.: Simulated annealing based fuzzy Markov game energy management for smart grids. In: Control and Decision Conference (CCDC), 2013 25th Chinese. IEEE, pp. 467–472, 2013.

[ZBG12] Zong, Yi; Bindner, Henrik W; Gehrke, Oliver: Real-time application of Different Predictive Control Strategies for Distributed Energy Resources Combining Demand Response and Wind. IEEE Transactions on Sustainable Energy, 2012.

[Zh13] Zhao, Weiliang et al.: Cost-Driven Residential Energy Management for Adaption of Smart Grid and Local Power Generation. 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

An Approach for a Cloud-based Contribution Margin Dashboard in the Field of Electricity Trading

Oliver Norkus[1], Brian D. Clark[2], Florian Merkel[3], Björn Friedrich[4], Jürgen Sauer[5], H.-Jürgen Appelrath[6]

Abstract: From our ongoing project, we present the collected requirements and business drivers from the energy industry imposed upon analytical information systems, specifically in the case of a contribution margin control as part of electricity trading. As a solution, we introduce our architecture for business intelligence in the cloud and apply these in the afore mentioned use case in the form of a research prototype. We provide the current implementation status at an early stage, on one hand to increase the transparency of the still young field of business intelligence in the cloud, and on the other hand to demonstrate the development potential of the technology for the energy sector.

Keywords: Cloud Computing, Business Intelligence, Electricity Trading, Cost Analysis, On Demand, Cloud-based Service.

1 Introduction

In our industrialized and increasingly globalized world, energy plays a major role. Without energy in the forms of uranium, coal, electricity, gas, and oil, our world and our daily life would look very different. The availability of light, heat, information, and power for the machinery locomotion is nowadays an inseparable part of our technology-society. We are in the midst of global competition for scarce resources. The energy supply is one of the central themes in science, politics, and business [AKM12].

The energy industry is undergoing changes, based on the increased production from renewable energy, the intensification of competition (especially in the context of liberalization and decentralized energy production) as well as a growing volatility, for example, in electricity trading. Electricity trading is a complex process in the contexts of market changes, legal situations, and further influence donors. In particular, the diversity

[1] University of Oldenburg, Department of Computer Science, Escherweg 2, 26121 Oldenburg, oliver.norkus@uni-oldenburg.de
[2] University of Oldenburg, brian.clark@uni-oldenburg.de
[3] University of Oldenburg, florian.merkel@uni-oldenburg.de
[4] University of Oldenburg, bjoern.friedrich@uni-oldenburg.de
[5] University of Oldenburg, Department of Computer Science, Ammerländer Heerstr. 114-118, 26121 Oldenburg, juergen.sauer@uni-oldenburg.de
[6] University of Oldenburg, Department of Computer Science, Escherweg 2, 26121 Oldenburg, appelrath@informatik.uni-oldenburg.de

of actors, the volatility of the relevant ratios, and the diversity of involved information technology (IT) systems all escalate the complexity [AKM12].

The relevant actors and roles have particularly increased by the unbundling of complexity. Through the direct marketing of electricity, remote meter reading, changes to smart grids, and (last but not least) through the trading of electricity capacity at the European Energy Exchange (EEX) raises the volatility and the velocity. Not only the minute or quarter hour accuracy, but also the volume of data results in new challenges for current IT systems. Through these factors, the challenges for business and statistical reporting as well as analysis becomes more convoluted, and as a result, the requirements on the used IT systems intensifies [NH14].

The functional requirements remain approximately the same in the areas of data procurement, processing, and analysis. The non-functional requirements increase in the areas of availability, accessibility, agility, query/analysis performance, and cost [No15].

Currently, the primarily used reporting and analysis systems are constructed monolithically. This means classical business intelligence (BI) solutions are complex, rigid, and expensive [BK10]. Therefore, traditional BI solutions are not suitable to cope with the increased demands. One approach to solving this problem is the use of cloud technology. Cloud computing has many properties that meet today's demands, especially in regards to analytical applications [No15].

Due to the novelty of the development in science and industry, there are not many field reports. Thus the acceptance of service offerings is still minor. Although first product versions from various manufactures are already available, they are in part incompatible, and cannot be combined or integrated with many existing systems, partly because fundamental issues have not been clarified in terms of business and software architecture [No15].

This is what we want to change with our approach, by utilizing a transparent, reusable cloud business intelligence system and the corresponding architecture applied in the form of a contribution margin control dashboard as part of the process of investigating the cost recovery of an energy supplier for new and existing customers in the context of direct marketing.

We aim to present the first approaches from our ongoing research project, and thus make our first experiences and results transparent. Therefore, we present in this paper required foundations (see section 2) of the requirements ascertained for cloud-based reporting and analysis systems and our use case for testing and evaluation (section 3), the approach for the architecture, and the current state of the implementation (section 4). Finally, we point out next steps and further work (section 5).

2 Foundations

The demand for increased flexibility in business is ever present. One aspect of this is the ability to enable employees to locate data and perform analyses without requiring intensive training of the complicated processes involved. At the same time, the commoditization of remote access to computing resources enables the utilization of powerful hardware from lightweight portable devices. The combination of these factors gives employees the potential to initiate and monitor a variety of complex processes in the middle of dynamic business situations, such as an on-site negotiation with prospective customers.

A solution to these needs is a datacenter containing powerful hardware for running analyses, combined with an intuitive web user interface designed for mobile devices. Using virtual machines (VM), additional hardware capacity can be allocated (or vacated) to follow with the current demand on system load. For the duration that additional capacity is needed, it can either be rented or reallocated from inactive hardware in a company's private data center. This process is invisible to the user, who can dispatch the analysis process from their portable device, and expect to have results shortly thereafter.

The concepts relating to supplying normal employees (without explicit system training) to performing analyses efficiently and effectively is covered in section 2.1, while technologies covering the ability to use commoditized hardware access is covered in section 2.2. The combination of these concepts is covered in section 2.3.

2.1 Business Intelligence

BI is a broad category of applications and technologies for gathering, storing, analyzing, and visualizing information. BI is associated with supporting IT resource management in order to optimize the process of decision-making, with a focus in providing decision makers with necessary information at just the right time [Ra09].

The understanding of the term *BI* ranges from multidimensional data structures on individual information systems to complex system landscapes analyzing large quantities of data for management. Analytical information systems focus on the provisioning of information and functional support for analyses in support of professionals and managers [Wa09].

2.2 Cloud Computing

Cloud Computing (CC) is defined as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [MG11].

A primary aspect of CC is the notion of a cloud service, defined as the provision of virtual IT resources (i.e. logical resources that are mapped to physical hardware) demonstrating the following characteristics [MG11], [VHH12]:

- *Resource pooling*: cloud services enable the shared utilization of physical resources by means of virtualization (virtual resources instead of physical hardware) and multi-tenancy (multi-tenant data management).
- *Rapid elasticity*: services can be immediately provisioned and released as demanded and the resources available for provisioning often appear to be virtually unlimited.
- *Measured service*: service utilization is measured and commonly monetized with a pay-per-use pricing model.
- *Broad network access*: cloud services are supported by ubiquitous network access and technical standardization.
- *On-demand self-service*: consumers can unilaterally provide themselves with services as they need them, due to extensive service automation on the side of the provider.

2.3 Business Intelligence Cloud

An early definition made in 2010 is from [SB10], which describes Business Intelligence Cloud (BI Cloud) as an IT architecture with the purpose of providing analytical capabilities as a service. A distinction can be made between the focus on the application (private, public, and hybrid) and for the architectural level (infrastructure, platform, and software).

More specifically in 2014, [NA14] states that analytical applications can be deployed as a cloud service, and that the outcome is called Business Intelligence as a Service (BIaaS). Such BI Cloud services reflect on the characteristics of typical cloud services features described below [NA15].

BIaaS enables organizations to flexibly respond to changing demands. Through flexible allocation, resources can be distributed depending upon project needs, especially to enable complex queries in an efficient manner. BIaaS analyses and reports can be flexibly integrated and orchestrated into processes.

The average employee (without specialized training) is able to automatically allocate and utilize resources around the particular services they are interested in through the system, without having to specifically invoke requests through the IT department, and be blocked waiting on a response from a human. Analytical applications in the cloud promote availability and reliability. They are accessible via an Internet connection from anywhere with only a web-enabled device.

BIaaS provides a high level of agility: reports and analyses can be adapted to directives and the needs of stakeholders. BIaaS not only provides flexibility at the user interface level with visualization and modeling, but also at the technical level, e.g., modifying cubes, extract, transform and load (ETL) processes, or multidimensional databases, which can all be processed in a simple and flexible manner. This is especially important when there is a demand to rapidly change operational IT landscapes (e.g. due to corporate transformations).

Analytical applications in the cloud promise very strong performance levels for queries and analyses, especially rapid deployments of complex reports and a short turnaround time when changing parameters. BIaaS can be billed on a per-use basis. Payment can be calculated for some aspect of the report, for example, the number of invocations, the hardware resources used (e.g. processing time), or access to proprietary data sets.

3 Requirements and Business Drivers

We performed face-to-face expert interviews from the energy industry for the discovery of requirements and business drivers. In accordance with agile software development processes, we performed requirements engineering in multiple iterations. We evaluated each interview to identify requirements and business drivers. Afterwards we framed the requirements utilizing techniques for requirements engineering [Ru09]. In the next step, the experts evaluated the requirements. Based upon the evaluation, the requirements were overhauled. The following interview started with a short presentation of the latest version. In the first interview, only basic requirements were discussed. In each of the subsequent interviews, the level of detail increased. After all experts agreed with the collected requirements, we considered the requirements from the viewpoint of a computer scientist. The interviews revealed a few use cases. The primary use case considered here is the contribution margin control.

Energy suppliers are chiefly interested in the costs a customer needs to pay per kilowatt-hour so that they realize a profit. To date, this calculation is mostly handcrafted with the help of spreadsheet software. This calculation is called *breakeven analysis* or *contribution margin control* and the result is the *cumulative cost*. The calculation of the cumulative costs is needed when energy suppliers make offers to prospective customers. The cumulative cost per kilowatt-hour is the essential part of any offering. When a contract between a customer and an energy supplier is up for extension, the current cumulative cost must be analyzed to calculate the cost for the new offer. The process of calculation lasts up to one week and has to be completed before a sales pitch takes place. This is especially disadvantageous for the field managers. If they visit a customer site for a customer pitch, it would be impossible for them to calculate a new value of the cumulative cost when necessary. They have to return to the office and find a new appointment with the customer. This is time-consuming for both sides.

During the expert interviews, we were able to derive the following requirements for a

contribution margin dashboard. With respect to building a prototype the fulfilled and here shown requirements are focused on functional aspects.

Availability: system redundancy is required to guarantee high availability, e.g. data backups and redundant infrastructure.

Accessibility: field managers are often on site for customer pitches. Thus, the service should be available on a wide range of devices, (not only on traditional computers, but also on mobile devices), and must be available over the Internet.

Accounting: given that BI systems are expensive to set up and maintain, it is necessary for clients to pay for their usage. The charges can be calculated either with respect to the complexity of the calculations or according to the volume of reports generated.

Performance: the duration of calculation, and data transport to the user should be minimized. Poor performance leads to decreased user acceptance of BI systems [NA15]. While the quality of mobile Internet connections an unavoidable factor, keeping transmitted data small and presenting overview before detail (like loading a graphic file over a dialup connection) helps to mitigate this.

Modularity: there are multiple processes related to the calculation of analyses. There is the potential for modifications or integrations of further analyses. Modularity enables the effective management of the system.

Expandability: many data sources are required to perform the analysis. The integration of additional data sources by users must be possible.

User empowerment: field managers should not need to write complex database queries or analyze raw data. The user interface needs be usable without special training, and allow, the user to intuitively understand the interface and the analysis result. This is key to user acceptance.

Web standards: nearly every mobile device or computer supports web technologies like HTML and JavaScript. The best way to bring the service to a wide variety of devices is to use the accepted standards.

The requirements *availability* and *accessibility* lead to the two additional computer science related requirements *scalability* and *flexibility*. Many field managers from multiple energy suppliers will use the system. The hardware resources have to be provisioned as required. If one server is overburdened, another server must be started to take the load away from the overburdened one. Through resource scaling, the service's full quality is available to every user.

Summarized, the service has to be available over many devices; field managers are often at customer sites and not at their office. The first aspect leads to flexibility according to the used devices and the second one to flexibility according to the access to the service. Both can be managed by providing the service over the Internet. Nearly every device in

the information technology domain supports web technologies. Therefore, providing the service over the Internet is preferable. Additionally the service could be used over the mobile web. This makes the service available over a great territory.

Scalability and flexibility are most relevant to the architectural. We show how we approach them in section 4 and how we intend to fulfill them in section 5.

4 Architectural Approach

In order to satisfy the identified requirements we envisioned the *Business Intelligence in the Cloud for Energy* (BICE) as a cloud service system, providing contribution margin analysis for electric utilities. As such, it needs to scale for large numbers of users while still responding quickly to requests. Additionally, the system has to enable users to add data sources and customize the generated reports. The analytical functionality is restricted to using preexisting data and does not incorporate any production planning.

In this section, we present our approach for a cloud-based contribution margin dashboard. In section 4.1, we describe the server portion and necessary system components that were identified in order to provide the required functionality. Following in section 4.2 is the description of our prototype, including the technology platform, data structures, and exemplary interaction between the components of the BICE system.

4.1 Components

The BICE system comprises of several components (seen in fig. 1) modelled with Unified Model Language (UML).

The user interacts directly with the Local Client component. It represents the user interface and control logic, as well as display data to the user. It uses the model-view-controller design pattern to manage complexity by separating the view from business logic. This component sends requests to the Load Balancer component, a proxy server that forwards the request to an available Dispatcher component running inside a VM. A Resource Monitor measures and reports the load on each VM, so that the Resource Management component can start up or shut down VM instances as needed. This response to the incoming load maintains Quality of Service for availability, accessibility, and query performance.

Depending on the type of the request, the Dispatcher communicates with the various other server components in order to answer the user request.

The User Database component stores relevant user data, such as username, password, permissions, the user's company, and usage statistics, and is encapsulated in the User Manager component. The Authorization and Rights Management component uses permissions to determine if a user may perform actions.

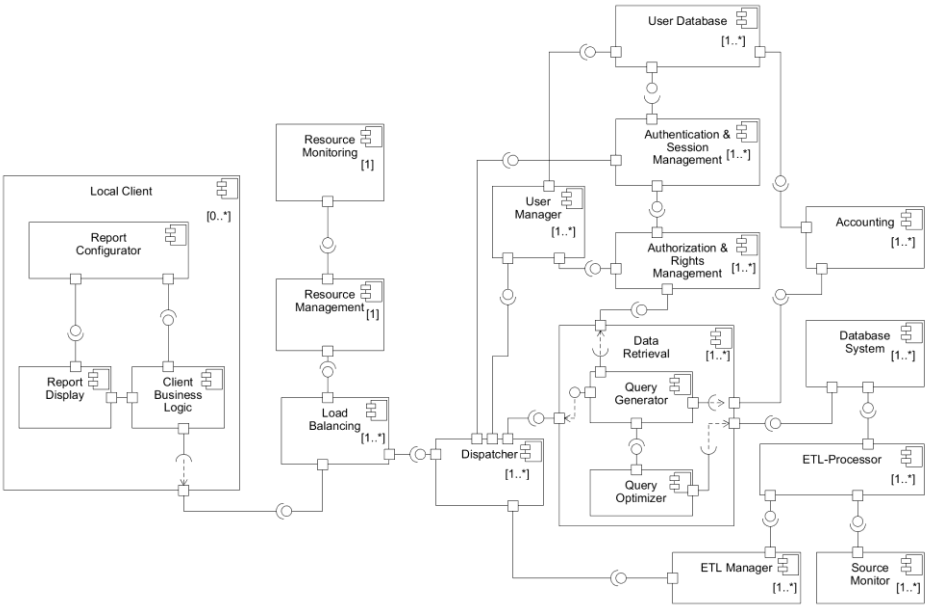


Fig. 1: Component diagram

Usage based billing is made possible by tracking the usage of individual users through the Accounting component and aggregating the data per electrical utility company. The user management related components will be deployed on a separate machine than the dispatcher.

In order to encapsulate and optimize database access, all analysis requests go through the Data Retrieval component running on separate VMs, so that resource intensive operations do not affect the overall responsiveness. To avoid unnecessary database queries and to enforce security, the Data Retrieval component verifies that the user who initiated a request has sufficient permissions to execute it.

The ETL-Manager, ETL-Processor, Source Monitor, and Database System components provide data extraction, transformation, storage, and analysis. The first three components play an important role in realizing flexible operation by enabling the user to incorporate data sources through the ETL-Manager component, which will load and transform data using an ETL-Processor. The Source Monitor component permits the monitoring of external data sources for changes, and signaling the corresponding ETL-Processor to start its work. During this process, heterogeneous data needs to be transformed into a format that facilitates the required data analyses provided by the system.

4.2 Prototype

We are developing a prototype of our architecture based on state of the art technologies provided by the Hasso-Plattner-Institute (HPI) to assess the feasibility of the conceptualization. The VMs required for running the components will be run and managed using HP Converged Cloud™, which adheres to the OpenStack™ architecture. The in-memory database system is a shared SAP HANA™ instance. The prototype will be evaluated in the scenario of analyzing the contribution margin of an electric utility companies for new and existing customers in the context of over the counter energy trading.

During the expert interviews, four key performance indicators were identified: total cost for supplying electricity to a customer, production cost, labor cost, and profit. These figures can be aggregated over time, the customer, the generators, and the distribution grids.

The data structure shown in fig. 2 was devised according to these dependencies. Despite the analytical focus of the system, the data is not structured as a star or snowflake scheme in order to avoid redundancy in an in-memory database system.

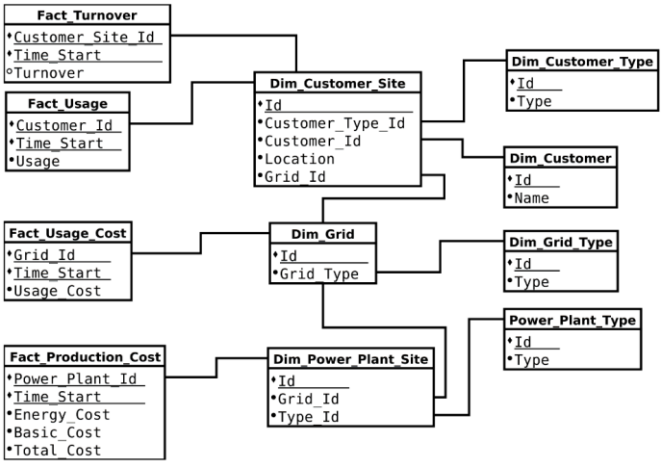


Fig. 2: Data fact table

The left direction in fig. 2 indicates the data fact tables that hold the data used to compute the indicators. Each record is linked to the entry in the dimension table it depends on, which in turn can reference additional records that provide information orthogonal to the dimension (e.g. Dim_Customer_Type of Dim_Customer_Site) and records that represent coarser representations along the dimension (e.g. Dim_Customer of Dim_Customer_site). The Time_Start field incorporates the time dimension in the fact tables. It describes the start of a 15-minute time interval for which the record is valid.

The Database System component stores data with the described data structure. The next section shows how, the system passes requests and responses back and forth.

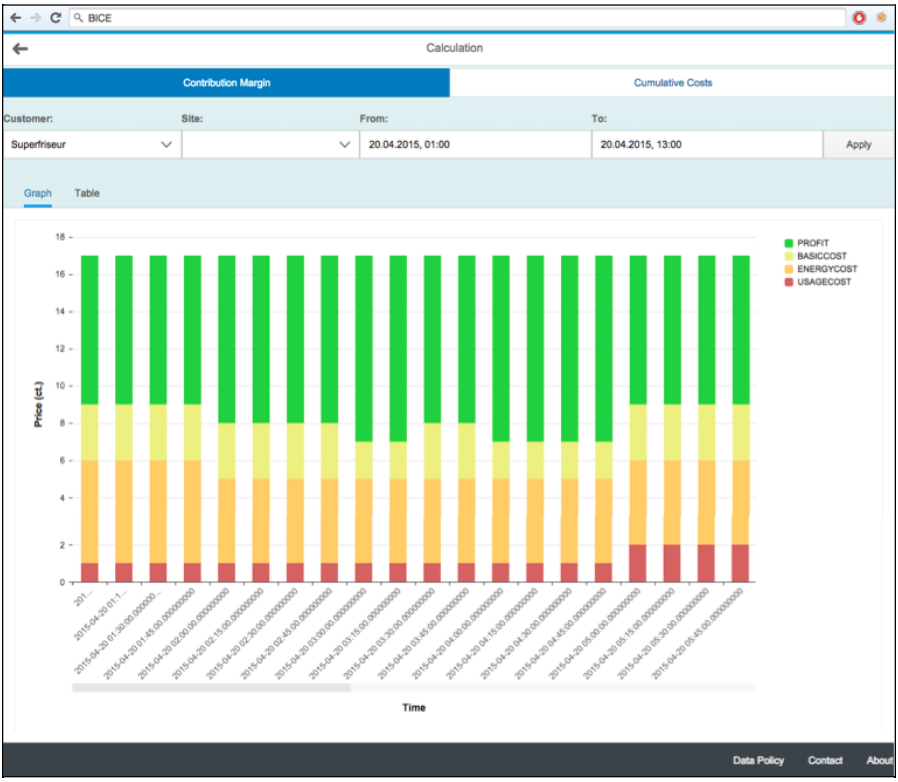


Fig. 3: Web interface for the dashboard

The user interacts with a web graphical user interface (GUI) like the one shown in fig. 3 to request and view data. The GUI targets mobile devices, specifically tablets. The choice of designing the GUI primarily for tablets is due to the larger screen size compared to smart phones, which can display more information at once, aiding in providing an overview of the data.

When the user is interacting with the system, they first need to log in with a username and password (not shown here). After setting the input fields to the desired values reflecting the user's request the *Apply* button sends the request to the server, which handles the necessary backend tasks and returns the response data. The client provides two views for displaying the requested information: a graph view, which shows the relevant indicators in a chart, and a view presenting information in table form.

In addition, the GUI also provides functionality for adding data sources and managing the user's own account.

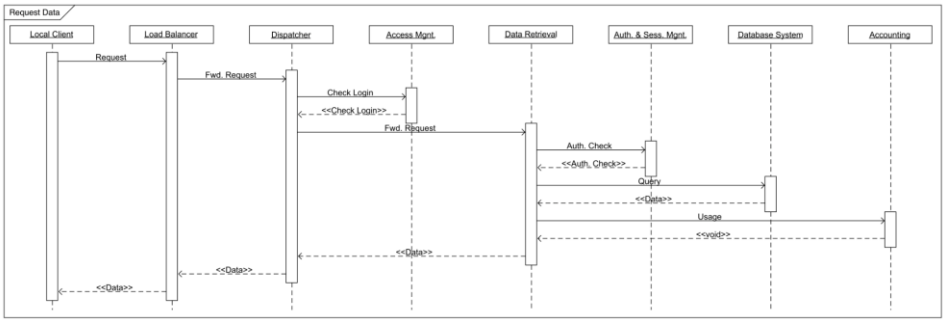


Fig. 4: Sequence diagram

The diagram in fig. 4 shows a typical sequence of interactions between the components as envisioned in the current architecture. It assumes the Local Client component is authenticated with the system when it sends a request for data. The request is received by the Load Balancer and passed to a Dispatcher component, which verifies that an authenticated client sent the request. The Dispatcher then asks for the respective data from the Data Retrieval component. After confirming that the Local Client instance has the proper access rights, it sends the actual database query. Additionally, resource usage is logged: possible measurements include the processing load on the database and/or the Data Retrieval component, or the time the request takes to be completed. The appropriate value is passed to the Accounting component, which tracks usage for individual users and provides functionality to aggregate the usage over all users of one electric utility company. After the database query returns, the requested data is passed back all the way to the client component while being transformed and converted as necessary.

5 Conclusion and Further Challenges

New challenges for BI arise with the velocity, volatility, and complexity of the markets. Old-fashioned BI systems cannot satisfy the new requirements. We introduced a flexible BI cloud system, which fulfills the new requirements.

In the use case, the BI cloud system performs the breakeven analysis faster and is (especially the field managers) more flexible. That must improve their negotiations and hence their business. However, there are many more processes in the energy industry. The breakeven analysis is only a single aspect. To date, many systems are used for providing a single process. We strive for a future of a single integrated system providing all processes in the energy business sector.

In this paper, we have shown the demands energy suppliers have for BICE, and how we gathered them. Based on the requirements, we developed the architecture for a prototype.

We finished developing the architecture for the BI cloud system and the HPI provides the hardware resources for the prototype. We are implementing based on the HP Converged Cloud™ and the SAP HANA™ in-memory database system. In respect to agile development strategies, the progress is will be presented to the experts throughout the development period. With their feedback, we are able to adapt the prototype closer to their needs. Moreover, the development and the improvement of existing features and components continues.

The next steps are to implement the further components shown in the architecture and deploy them to the mentioned hardware. The prototype is going to be evaluated with regards to the use case. We will present the results of our evaluation and further findings by means of a report and a poster as part of the HPI Future SOC Lab Days on the 4th November 2015.

Literaturverzeichnis

[AKM12] Appelrath, H.-J., Kagermann, H., Mayer, C. (Eds.). Future Energy Grid: Migrationspfade ins Internet der Energie. Springer-Verlag, 2012.

[NH15] Norkus, O.: Haas, M.: Stromhandel in der Cloud. ew - Magazin für die Energiewirtschaft, EW Medien und Kongresse, Frankfurt, S.44-46, 4/2015

[No15] Norkus, O.: Ein Ansatz zur Standardisierung von BI in der Cloud. In (Uwe Aßmann, Birgit Demuth, Thorsten Spitta, Georg Püschel, Ronny Kaiser): Proceedings of Software Engineering & Management 2015, Dresden, Gesellschaft für Informatik, Bonn, Bonner Köllen Verlag, S.299-303, LNI, P-239, 3/2015

[Ru09] Rupp, C.: Requirementsengineering und -Management. Hanser Verlag, 5. Auflage, 2009.

[BK10] Baars, H.: Kemper, H-G.: Business intelligence in the cloud? PACIS 2010 Proceedings. Paper 145. 2010.

[Ra09] Ranjann, J.: Business Intelligence: Concepts, Components, Techniques and Benefits. Journal of Theoretical and Applied Information Technology.Vol 9. No 1. (pp 060-070)

[Wa09] Watson, H. J.: Tutorial Business Intelligence - Past, Present, and Future. Communications of the AIS 25 (2009, 1), pp. 487-510.

[MG11] Mell, P.; Grance, T.: The NIST definition of cloud computing. 2011. National Institute of Standards and Technology, 2011

[VHH12] Vossen, G.; Haselmann, T.; Hoeren, T.: Cloud-Computing für Unternehmen. Technische, wirtschaftliche, rechtliche und organisatorische Aspekte. dpunkt, Heidelberg (2012).

[SB10] Seufert, A.; Bernhardt, N:. Business Intelligence und Cloud Computing. HMD Praxis der Wirtschaftsinformatik, 47(5):34–41, 2010.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Auswertung von Energiedaten – Ein Ansatz zur Kombination einer soziologischen Perspektive mit der Methodik des Data Minings

Tobias Weiß[1] und Marco Krause[1]

Abstract: Die Methoden des Data Minings liefern sinnvolle Ansätze zur Weiterverarbeitung und zielgruppengenauen Segmentierung von Kunden für innovative Marketingansätze. Dies wird bei Energieversorgern durch zunehmenden Wettbewerb und neuen Marktbegleitern ein immer wichtigeres Thema. Bestehende Forschungen beschäftigen sich mit der Segmentierung von Lastgängen nach maximaler Ähnlichkeit des Profils. Problematisch ist dabei bislang die Integration von kundenspezifischen Daten sowie deren Erhebung im Rahmen einer Lernstichprobe. Die vorliegende Publikation widmet sich dieser Thematik und adressiert die Kombination von soziodemografischen und lebensstilspezifischen Daten mit der aktuell verbreiteten Anwendung des Data Minings. Die Untersuchung nimmt dabei die Integration einer konsumsoziologischen Perspektive vor, mit dem Ziel die wichtigsten Faktoren auf Grundlage relevanter Literatur theoretisch herzuleiten. Diese Faktoren werden dann mit der Methodik des Data Minings verknüpft, um einen Ansatz zu generieren, welcher ein umfassenderes Bild des Konsumenten liefert und so dem Marketing eine breitere Basis für die zielgruppenspezifische und gleichzeitig datengetriebene Ansprache zur Verfügung stellt.

Keywords: Energy Analytics, Data Mining, Business Intelligence, Smart Metering, milieuspezifischer Energiekonsum, soziodemografische Faktoren, milieuspezifische Konsummuster, Lebensstil und Energie

1 Einleitung

Der deutsche Energiemarkt befindet sich im Wandel. Viele dieser neuartigen Bewegungen in diesem traditionellen Marktumfeld sind durch politische Eingriffe verursacht. Im Rahmen der Liberalisierung, eingeleitet 1999 durch europäische Richtlinien, wurde der Markt aufgebrochen und Unbundling etabliert. Auch aktuell beeinflussen neue Rahmenbedingungen das Marktgefüge, beispielsweise die Richtlinien im Rahmen der Energiewende. Diese fordert aus regulatorischer Perspektive nicht nur beträchtliche Einsparungen, sondern ebenfalls den Ausbau der Stromnetze und den Rollout von Smart Metern, als intelligente und fernauslesbare, digitale Evolution der Stromzähler. Sollten bislang lt. EnWG Haushalte mit einem Jahresbedarf über 6.000 kwH sowie Neubauten und Renovierungen mit Smart Metern ausgestattet werden (vgl. [BJ05a]), zeigt das aktuelle Eckpunktpapier der Bundesregierung eine detailliertere

[1] Technische Universität Dresden, Lehrstuhl für Wirtschaftsinformatik – Business Intelligence Research, Münchner Platz 3, 01187 Dresden, tobias.weiss@tu-dresden.de

Sicht. So sollen, basierend auf der durchgeführten Kosten-Nutzen-Analyse des BMWi (vgl. [EY13]), Haushalte mit einem Jahresbedarf unter 6.000 kWh grundsätzlich ausgenommen, sowie ein schrittweiser Rollout bis ins Jahr 2021 vorgenommen werden (vgl. [BWE15]). Auch hinsichtlich der eingesetzten Smart-Meter-Hardware sind klare Richtlinien zu beachten, maßgeblich beeinflusst durch das BSI [vlg. BSI SM Richtlinie].

Ein weiterer Punkt, der bereits seit der Liberalisierung sichtbar ist: der Markt an sich wird ebenso dynamischer. Neue Unternehmen mit innovativeren Geschäftsmodellen, als jene die die klassischen Energieversorger bieten, betreten den Markt und sorgen für eine Neuverteilung. So sehen sich Energieversorger einem stets steigendem Wettbewerb und einer rückläufigen Nachfrage ausgesetzt (vgl. [Se12], S. 82-84). Bislang konnte man dies bei neuen Versorgern betrachten, die oftmals weder über Netz oder Produktion verfügen, und auf Zielgruppen mit einem maßgeschneiderten Angebot zugehen (bspw. ökologische Energieerzeugung aus nachhaltigen Quellen, für bewusste Zielgruppen). Das Angebot ist rational betrachtet aber identisch: es handelt sich um Strom, ein unemotionales Commodity-Gut. Dieser Tatbestand resultiert daraus, dass die meisten Verbraucher nur einmal im Jahr Kontakt mit Ihrem Energieversorger haben. Leider meist im negativen Sinne, denn ein Mal pro Jahr erfolgt die Rechnungsstellung, evtl. mit einer Nachzahlung. Dabei haben klassische Energieversorger mit einem großen Kundenbestand aus Zeiten vor der Liberalisierung einen großen Vorteil: sie besetzen die Kundenschnittstelle. Mit dieser Marktmacht und letztendlich auch finanziellen Möglichkeiten haben sie die Chance so viele Kunden wie möglich zu halten und neue hinzuzugewinnen. Im Gegenzug haben innovative Neulinge die Chance sich in Nischen zu behaupten.

2 Potential für neue Geschäftsmodelle

Dazu bedarf es neue Angebote und Geschäftsmodelle, basierend auf innovativen Technologien, die heute schon verfügbar sind. Experten sind sich einig, dass das Abwarten und Festhalten an alten Strategien nicht der erfolgversprechendste Ansatz ist (vgl. [EY15], S. 7). Es wurde weiterhin festgestellt, dass Unternehmen, die einen klar vermittelbaren Mehrwert für Kunden bieten und sich gezielt auch auf neue Trends einlassen, in Zukunft eine deutlich bessere Perspektive auf dem Markt haben werden (vgl. [EY15], S. 9). Es lässt sich daher schlussfolgern, dass neue Geschäftsmodelle vor allem kundenzentriert konstruiert werden müssen. Die Frage ist, wie man dies in Kombination mit neuen Daten realisieren kann, um insbesondere der Marketingabteilung Mehrwerte zu erschließen.

Geschehen soll dies im ersten Schritt durch die Anwendung von Algorithmen des Data Minings zur Kundensegmentierung. Was in anderen Branchen, wie bei Banken zum Kreditscoring, aus rein finanziellen Gründen seit einiger Zeit ein fester Bestandteil der Kundenbearbeitung ist, ist bei Energieversorgungsunternehmen noch nicht stark verbreitet. Hier findet eine Kundensegmentierung und Aufbereitung für das Marketing nach klassischen Ansätzen statt: an eine Auswahl an Kunden werden Umfragen

versendet, auf Basis dieser Daten wird dann eine Segmentierung durchgeführt. Bei dieser sind sowohl Persönlichkeitsmerkmale als auch Produktmerkmale enthalten (vgl. [Mo08]). Heraus kommt dabei bspw. eine Charakteristik für ökologisch besonders nachhaltig agierende Kunden.

Die Nachteile dieses Vorgehens liegen jedoch auf der Hand. In der Regel ist die Antwortquote einer solchen Umfrage gering, bei ca. 5%. Des Weiteren ist diese sehr zeitaufwendig und damit letztendlich kostenintensiv sowie schlecht skalierbar (vgl. [AR13], S.9). Sinnvoller wäre es, wenn diese Kundensegmentierung und Ableitung von Segmentmerkmalen direkt aus den zukünftig entstehenden, großen Datenmengen des Smart Meterings extrahiert werden könnten. Basierend auf Algorithmen ist diese Segmentierung mit den Verbrauchsdaten eines jeden Kunden durchführbar, damit nahezu beliebig skalierbar und in kurzer Zeit mit einem geringen Kostenaufwand durchführbar. Die Kundensegmente würden dann, nach einer initialen Lernphase sowie der Integration von Expertenwissen, vorrangig Datengetrieben gebildet werden.

3 Forschungsdesign

Ziel der vorliegenden Publikation ist die Erstellung eines Ansatzes zur datengetriebenen Kundensegmentierung. Dabei sollen die soziodemografischen Merkmale und Werte der Konsumenten in die Methodik des Data Minings integriert werden. Mittels eines interdisziplinären Zugangs werden soziologische Elemente sowie Komponenten der Wirtschaftsinformatik miteinander verbunden, um die Perspektive auf den Konsumenten zu erweitern und den Erkenntnisgewinn zu steigern. Der Schwerpunkt liegt dabei in der Identifikation von Determinanten auf den Energieverbrauch von Privatkunden. Dabei sollen Merkmale mit einem hohen Einfluss auf das Verbrauchsverhalten abgeleitet werden. Die Stadtwerkestudie unterstreicht diese Idee deutlich mit empirisch erhobenen Aussagen. So werden sich in den nächsten zwei bis drei Jahren 73% der befragten Stadtwerke mit dem Themenbereich „Absatz/Marketing/Kundenbetreuung/CRM“ stark oder sehr stark auseinandersetzen (vgl. [EY15], S. 14).

Um die gestellte Forschungsfrage adäquat beantworten zu können, findet für die Sicht der Wirtschaftsinformatik zur Vorbereitung die Forschungsmethode des Reviews Anwendung. Dieser Ansatz erstreckt sich von der systematischen Problemformulierung bis hin zur Auswertung und ist orientiert an dem Vorgehen nach Fettke (vgl. [Fe06]). Ziel dieses Reviews ist die Identifikation eines Vorgehens zur datengetriebenen Segmentierung. Im weiteren Verlauf erfolgt eine prototypische Implementierung. Auch aus soziologischer Sicht fußt die Bearbeitung der Forschungsfrage auf der Methode Literaturanalyse. Die Grundlage bilden dabei Publikationen, welche die Themenbereiche von Konsummustern im Zusammenhang mit Milieuzugehörigkeit und soziodemografischen Faktoren umfassen. Das Ziel ist die Herausarbeitung und systematische Darstellung von soziodemografischen Faktoren und Lebensstilaspekten, welche einen Einfluss auf das Konsumverhalten im Bereich Energie ausüben.

4 Status Quo

Bei all den Überlegungen zur Segmentierung der Kunden anhand ihres Verbrauchsverhaltens muss man sich vor Augen halten, dass es eine derartige Unterteilung der Verbraucher bereits seit Langem gibt. Standardlastprofile (SLP) sorgen bei Kunden, die nicht über eine registrierende Leistungsmessung (RLP) tarifiert werden (ab 100.000 KWh, vgl. [BJ05b], §12), für eine Grundlage zur Prognose des Strombedarfs im Laufe der Zeit. Mit besonderem Fokus auf Privathaushalte betrachten wir insbesondere das Profil „H0“. Natürlich ist davon auszugehen, dass das reale Verbrauchsverhalten unterschiedlicher Haushalte nicht immer dem standardisierten Lastgang des H0-Profils folgen wird. Die Frage ist jedoch, wie viele Kunden eines Energieversorgers sich tatsächlich korrekt mittels H0-Profile prognostizieren lassen.

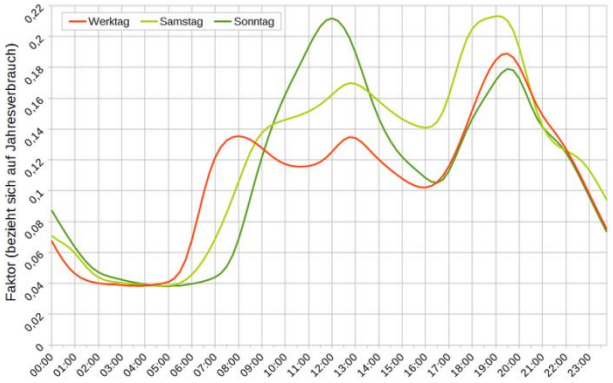


Abb. 1: H0-Standardlastprofil nach VDEW, Winterhalbjahr (vgl. [VDE15], [Wi14])

Je mehr Kunden ein zum SLP ähnliches Verbrauchsverhalten aufweisen, desto zielgerichteter können Lastprognosen der Energieversorger optimiert werden. Was bei dieser Betrachtungsweise jedoch fehlt, ist eine Integration von konsumentenspezifischen Faktoren der Kunden. Aus Marketinggesichtspunkten ist jeder Haushalt innerhalb der H0-Gruppe völlig identisch, lediglich unterschieden durch Werk- Sonn- und Samstage sowie Jahreszeiten (Sommer, Winter, Übergangszeit). Dies erschwert die Kundensegmentierung. Eine genauere Unterscheidung ist oftmals nur durch beschriebene Umfragen-Methodik zur Datenerhebung möglich.

Für die Anwendung von Methoden des Data Minings konnten mittels eines Reviews der Literatur gem. [Fe06] empfohlene Vorgehensweisen und Algorithmen identifiziert werden, welche sich insbesondere auch für die Anwendung auf Energiedaten, respektive Zeitreihen, eignen. Die 25 relevanten Publikationen prüfen entweder dedizierte Algorithmen hinsichtlich der Anwendbarkeit für die Segmentierung von Lastgängen bzw. Zeitreihendaten im Energiekontext, oder beschreiben den Gesamtprozess zur Segmentierung (vgl. [Wa14]). Diese sind breiter angelegt und betrachten, um einen konkreten Ausschnitt hervorzuheben, auch Besonderheiten bei der Vorverarbeitung von

Energiedaten. So können bspw. Ausreißer oder Missing Values die Datenqualität und damit das Analyseergebnis maßgeblich beeinflussen, und müssen daher mit geeigneten Methoden vorbehandelt werden.

Da dieser Review nicht im Vordergrund der vorliegenden Publikation stehen soll, werden an dieser Stelle Publikationen gezielt herausgehoben, die exemplarisch für die Ergebnisse stehen und dem Leser einen guten weiterführenden Einblick bieten. Ein detaillierter Vergleich der Algorithmen ist vielmehr für die nachfolgende prototypische Implementierung der im folgenden Hauptteil hergeleiteten Erkenntnisse über die Eignung von verschiedenen soziodemografischen Faktoren von Relevanz. Ein geeignetes Vorgehen zur Auswertung von Smart-Meter-Daten präsentieren [FRV05] in ihrer Publikation zur Charakterisierung von Stromkunden, basierend auf klassischen Methoden des Data Minings (wie bspw. KDD). Einen Überblick über Segmentierungs-Methoden, die innerhalb eines solchen Vorgehens Anwendung finden, präsentiert [Ch12]. Auch er nimmt zunächst eine Abbildung des Gesamtprozesses vor (vgl. [Ch12], S. 69) um danach Methoden zu Benchmarken. Auch hier zeigt sich, dass Methoden mit einer Ausreißer-Behandlung bessere Analyseergebnisse erzielen.

5 Auswertung von Energiedaten mit Methoden des Data Minings

5.1 Vorgehen der protypischen Umsetzung

Es hat sich im Rahmen des Reviews gezeigt, dass sowohl eine grundsätzliche Vorgehensweise existiert, als auch eine Auswahl von möglichen Algorithmen zur Datenanalyse. Die Entscheidung, welche Vorverarbeitungsschritte und welcher Algorithmus im konkreten Fall tatsächlich Anwendung findet, ist stark determiniert durch die Anwendungssituation und Datenlage. Dazu wurde im Beitrag eine prototypische Umsetzung des Vorgehensmodells vorgenommen, zusätzlich orientiert am, im Bereich des Data Minings geläufigen, Vorgehen nach CRISP-DM (vgl. [Ch00]).

Ziel der prototypischen Umsetzung, analog zum Business Understanding des Vorgehensmodells, war eine Segmentierung von Lastgängen eines Pilotprojektes in Hinblick auf möglichst ähnliche Lastgangprofile. Kunden mit einem ähnlichen Verbrauchsverhalten im Laufe des Tages sollten also demselben Segment zugeordnet werden. Bevor wir einen Blick auf das Gesamtvorgehen werfen, sollen zunächst die Eigenarten der Daten im Rahmen des Data Understandings genauer betrachtet werden.

5.2 Datenlage des Pilotprojektes

Eine nicht unwesentliche Herausforderung von Forschungen im Bereich der Energiedatenanalyse ist der Mangel an Daten aus Pilotprojekten, welche zur Optimierung und Evaluierung von Algorithmen, Methoden und Verfahren herangezogen

werden können. Wie auch in [AR13], S. 3, beschrieben, befindet sich die Energiedatenanalyse in den Kinderschuhen. Benannte Autoren stützen sich in ihrer Publikation, welche ebenfalls kundenspezifische Daten miteinbezieht, auf einen Datensatz von 1.100 Google-Mitarbeiter-Haushalten in den USA. Dabei liegen Zeitreihen des Stromverbrauchs im 10-Minuten-Intervall vor sowie konsumentenspezifische Daten, die im Rahmen einer Umfrage erhoben wurden (vgl. [AR13], S. 5). Nicht von der Hand zu weisen sind die Schwierigkeiten, welche bei der Analyse von Daten dieser Herkunft entstehen können. Google-Mitarbeiter repräsentieren nur einen bestimmten Ausschnitt der Bevölkerung, haben ein vermeintlich überdurchschnittliches Einkommensniveau, eine technologische Affinität und gut ausgestattete Haushalte. Eine kritische Betrachtung dieser Daten ist daher angebracht.

Unsere prototypische Umsetzung stützt sich auf Daten aus dem National Smart Meter Trial of the Electric Ireland and Sustainable Energy Authority of Ireland, der zwischen 2009 und 2010 durchgeführt wurde und über 5.000 irische Haushalte und KMUs beinhaltet. Das Intervall der Auslesung von Stromverbrauchsdaten lag bei 30 Minuten. Es handelt sich damit um eine der größten statistischen Erhebungen, die aktuell bekannt sind (vgl. [SEAI12]). Zusätzlich wurden konsumentenspezifische Daten mittels einer Umfrage erhoben, welche umfangreiche Informationen zu jedem Haushalt enthalten.

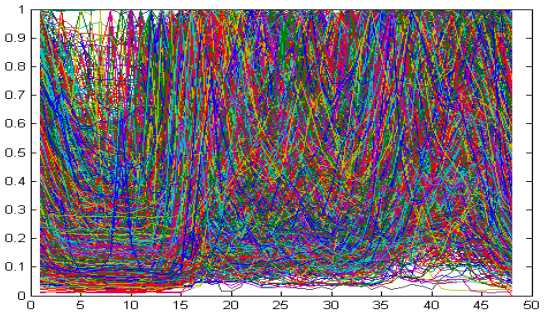


Abb. 2: Darstellung aller Lastgänge des Pilot-Datensatzes (vgl. [SEAI12], [Wa14], S. 53).

Bei einer initialen Visualisierung der enthaltenen Lastgänge in Abb. 2 wird auch direkt die eingangs motivierte und im Status Quo visualisierte Situation deutlich: es existiert in der Realität eben kein standardisierter Lastgang für Privathaushalte, wie der Standardlastgang H0 dies versucht. Selbstverständlich ist das auch nicht der Anspruch von H0, eine feinere Untergliederung kann jedoch der Prognosegüte sowie der Kundenansprache durch das Marketing nur hilfreich sein.

5.3 Vorgehensmodell der prototypischen Umsetzung

Im Folgenden ist das generelle Vorgehen schematisch abgebildet, angelehnt am identifizierten Vorgehen der Literatur, unter Beachtung der Datenlage:

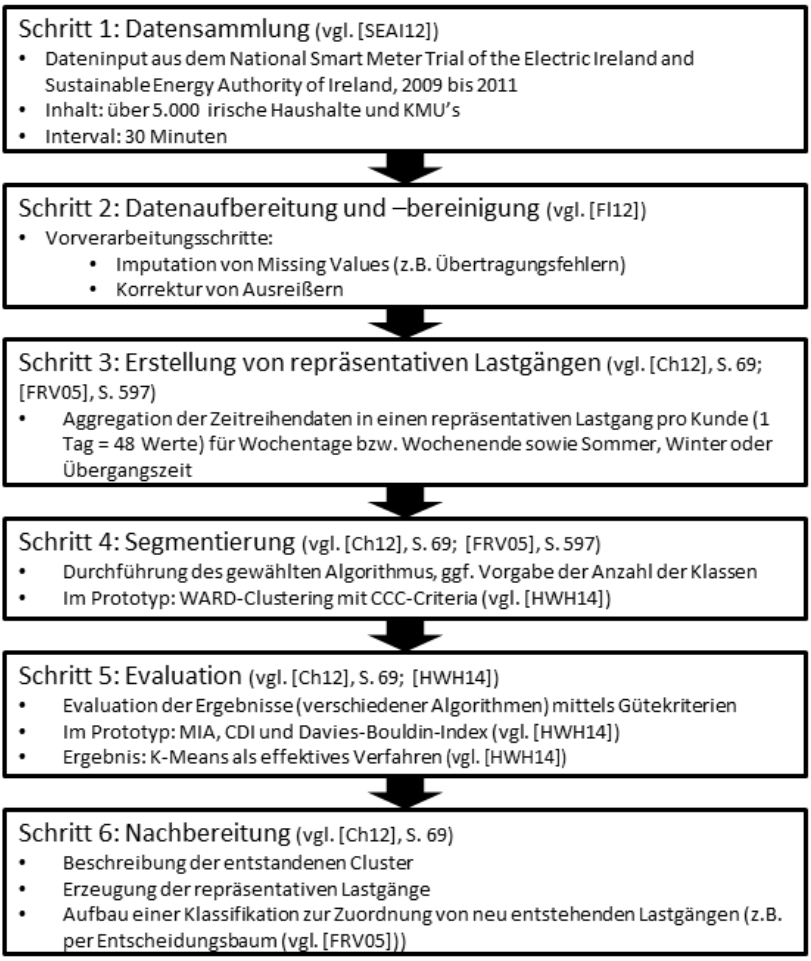


Abb. 3: Vorgehensmodell der prototypischen Implementierung (in Anlehnung an [Ch12], [FRV05] und [HWH14])

5.4 Ergebnis der prototypischen Umsetzung

Als Ergebnis der Segmentierung wurden 6 Cluster identifiziert, deren Lastgänge eine maximale Ähnlichkeit aufweisen. Wie in der nachfolgenden Grafik ersichtlich ist, gibt es typische „Häufungen" von repräsentativen Lastgängen (in unserem Fall konkrete Haushalte), aber auch zahlreiche Überlappungen, die möglicherweise nicht nur einen Einfluss auf den Verlauf des Profils haben, sondern auch auf dahinterliegende konsumentenspezifische Eigenschaften.

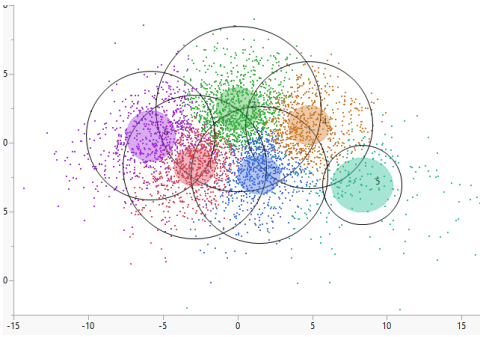
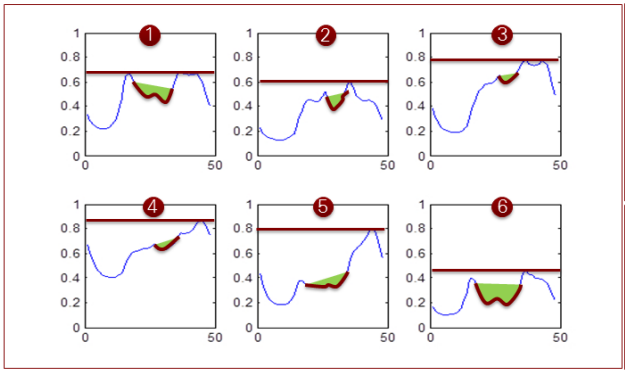


Abb. 4: Visualisierung der Cluster (vgl. [HWH14])

Die repräsentativen Lastgänge der einzelnen Cluster zeigen folgende Verteilung und Strukturen:



Cluster Summary

Cluster	Count
1	824
2	1091
3	750
4	648
5	251
6	661

1. High mornings
2. Normal profile
3. Normal high
4. High consumption
5. High evenings
6. Low consumption

Abb. 5: Repräsentative Lastgänge für jedes Cluster, mit Erläuterung und Anzahl (vgl. [HWH14])

Mit dieser Visualisierung lassen sich bereits rein optisch Hypothesen zu einem möglichen Tagesablauf oder sozialen Gewohnheiten der betroffenen Haushalte bilden, wie auch im rechten Kasten der Abbildung 5 in Form von Profilbeschreibungen ersichtlich ist. Um den Anforderungen der Zielgruppe Marketing jedoch gerecht zu werden, ist eine reine Segmentierung der Lastgänge nach Ähnlichkeit nicht hinreichend. Vielmehr sind die wichtigsten Charakteristiken der sozialen Gruppen, die sich hinter den repräsentativen Lastgängen befinden, interessant.

6 Einbettung von Lastgängen in Konsummuster

Unser Pilotdatensatz enthält dazu enorm viele Informationen über die Menschen hinter dem Energieverbrauch, wie bspw. Art des Arbeitsverhältnisses, wie viele Menschen im Haushalt leben, oder die Einstellung zu ökologischen Fragestellungen (vgl. [SEAI12]).

Die Frage ist nun, welche dieser erhobenen Eigenschaften wirklich erforderlich sind, um einen Bezug zwischen Lastgang und Kundengruppe herzustellen, und Marketingaktivitäten durch erhöhtes Kundenwissen zu unterstützen. Je konkreter im Vorfeld die Kenntnisse bezüglich der Thematik und vorliegender theoretischer Konzepte sind, desto effizienter kann das Instrument zur Erhebung und Messung dieser Faktoren konstruiert werden. So ist es von Bedeutung möglichst genau die konkreten Parameter festzulegen, welche gemessen werden sollen. Dabei gilt: je weniger Merkmale gemessen werden müssen, desto zeiteffizienter ist die Erstellung einer passenden Lernstichprobe zum Herstellen der Verbindung zwischen Lastgang und Haushaltstyp.

So sollen anknüpfend an eine konsumsoziologische Perspektive Hypothesen bezüglich der konsumentenspezifischen Determinanten des Energieverbrauchs abgeleitet werden. Diese konsumentenspezifischen Determinanten, welche auf den Energiekonsum rekurrieren, sind es im Weiteren, die in das bisherige Datenkonzept integriert werden sollen, um dadurch die Sicht und den Erkenntnisgewinn bezüglich der Konsumenten zu erweitern. Hierbei wird eine Beschränkung dahingehend vorgenommen, dass nur Elemente, welche sich auf den Konsum der Energie auswirken, in das Modell aufgenommen werden sollen, um dadurch die trennscharfe Kontur des Indikators zu wahren. Perspektivisch soll das Konsummuster des Haushalts nur anhand der Charakteristik des Lastgangs hergeleitet werden. Dieses datengetriebene Vorgehen eröffnet die Möglichkeit der Aufdeckung von Kundengruppen, auch ohne das Durchführen einer Befragung der Konsumenten.

6.1 Konsummuster und ihre konsumentenspezifische Einbettung

Konsumakte des Energieverbrauchs verlaufen für die meisten privaten Konsumenten oftmals unreflektiert. Solche Konsumhandlungen sind eingebettet in spezifische alltägliche Verhaltensmuster und Routinen (vgl. [Ch09], S. 7). Die Entwicklung von subjektiven Konsumpraktiken ist dabei einerseits durch gesellschaftlich aktuelle Diskurse und Trends, anderseits aber auch durch jeweils konsumentenspezifische Merkmale determiniert. Diese umfassen beispielsweise das Einkommen, den Grad der Bildung, das Geschlecht sowie milieuspezifische Werte und Einstellungen (vgl. [Ch09], S. 7). So ist auch der Energieverbrauch der Konsumenten nicht ausschließlich durch technische Aspekte bestimmt, sondern auch durch soziodemografische Faktoren sowie dem Lebensstil der handelnden Akteure (vgl. [GB11], S.2). Definiert werden Lebensstile an dieser Stelle als *„gruppenspezifische Formen der alltäglichen Lebensführung, -deutung und -symbolisierung von Individuen im Rahmen ökonomischer, politischer und sozialer Kontexte. In ihnen sind (sic!) die objektive Dimension sozialer Lagen mit der subjektiven Dimension von Mentalitäten und Wertvorstellungen verknüpft.“* (Sp10], S. 5 zitiert nach [Re99], S. 53) Entlang der Faktoren der soziodemografischen Merkmale und der Lebensstile lassen sich Einteilungen in eine Vielzahl verschiedener Milieus vornehmen. Demnach bilden Milieus Gebilde, deren Akteure sich durch einen bestimmten Ähnlichkeitsgrad hinsichtlich ihres Lebensstils und ihrer sozialen Lage kennzeichnen. Dementsprechend lässt sich den Mitgliedern eines Milieus auch eine

gewisse Homogenität hinsichtlich ihrer Konsumgewohnheiten unterstellen (vgl. [Ch09], S. 13-14). Dabei wird davon ausgegangen, dass der Konsum von den jeweils subjektiven Lebenszielen und Präferenzen des Konsumenten determiniert wird, welche sich wiederum in den jeweiligen Lebensstilen manifestieren (vgl. [Ca09], S. 19). Dieses Konstrukt entspricht den Vorstellungen eines spezifischen Clusters im Data Mining.

Demnach zeichnet sich an dieser Stelle ein komplementäres Verhältnis zwischen den konsumentenspezifischen Faktoren, also den soziodemografischen und milieuspezifischen Merkmalen und dem energiebezogenen Konsumstil ab. So konnten zum Beispiel [GB11] in ihrer empirischen Studie einen signifikanten Zusammenhang zwischen dem Lebensstil des Konsumenten und dessen Wahl bezüglich des Konsums eines bestimmten Heizungssystems nachweisen (vgl. [GB11], S. 5).

Dementsprechend erhalten solche Milieu- und Lebensstilansätze eine stetig wachsende Aufmerksamkeit, da sie es vermögen, ein umfassenderes Bild des Konsumenten und dessen Bewegungen auf dem Markt zu liefern. Mithilfe solcher Konzepte lassen sich zum Beispiel bezogen auf den Energiemarkt, verschiedene Energiekulturen identifizieren, auf welche durch eine spezifische Ansprache und Ausgestaltung der Vermarktung Einfluss genommen werden kann, um so einen optimalen Nachhaltigkeitsgewinn zu erzielen (vgl. [Sp10], S.5-6).

6.2 Relevante konsumentenspezifische Faktoren

Im Folgenden sollen nun die im Rahmen der Literaturanalyse abgeleiteten Faktoren der Konsumenten dargestellt und ins Verhältnis zueinander gesetzt werden. Die hier aufgelisteten Parameter sollen dabei als Ausgangspunkt dienen und den Anstoß für weitergehende Forschung hinsichtlich dieser Thematik fördern.

Verschiedene Studien haben die Thematik des Zusammenhangs zwischen soziodemografischen Merkmalen und der Einstellung bezüglich des Energiekonsums bearbeitet. Im Folgenden sollen die Studien von [NGB07], [Ca09], [Pr10] und [BGS12] als Ausgangspunkt zur Ableitung relevanter Konsumentenmerkmale bezüglich des Energiekonsums herangezogen werden.

Die genannten Studien identifizierten eine Reihe von soziodemografischen Merkmalen und konsumentenspezifischen Werten, die in je unterschiedlicher Ausprägung und Kombination einen Einfluss auf den Energiekonsum ausüben. Unter der Berücksichtigung der Hypothese, dass die soziodemografischen Faktoren und die spezifischen Ausprägungen verschiedener Aspekte der Lebensstile der Konsumenten in deren Energieverbrauch intervenieren, kann das Wissen um die Positionierung der Konsumenten im sozialen Raum weiterverarbeitet werden. Die Verschränkung dieser verschiedenen Parameter liefert in der Folge eine Basis, auf der konsumentenspezifische Typologien des Energiekonsums erstellt werden können (vgl. [Ca09], S. 27).

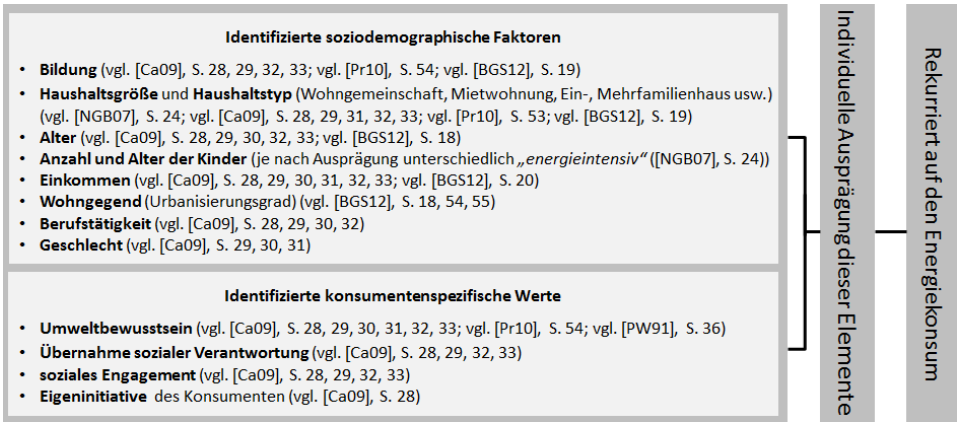


Abb. 6: Übersicht der identifizierten konsumentenspezifischen Merkmale

So identifizierte beispielsweise [Ca09] in ihrer Studie die sogenannten WELSKO (Wert-Lebensstil-Konsumverhalten) Haushaltstypen, welche durch ähnliche konsumentenspezifische Merkmale gekennzeichnet sind und in der Tendenz einen ähnlichen energiebezogenen Konsum aufzeigen. Als Beispiel soll an dieser Stelle der Haushaltstyp des **Alternativ-Umweltbewussten** umschrieben werden. Angehörige dieses Segments weisen ein durchschnittliches Alter von 30 Jahren auf, zudem überwiegt in dieser Gruppe der Anteil der Frauen. Die meisten Angehörigen dieses Typus weisen mindesten einen Realschulabschluss auf und gehören zur unteren bis mittleren Einkommensschicht. Die Wohnungssituation stellt sich dabei so dar, dass die Angehörigen meist in Mehrpersonenhaushalten zur Miete leben. Werte, die in dieser Gruppe von Bedeutung sind, sind die Verantwortung gegenüber den folgenden Generationen, der Umweltschutz sowie ein soziales und kulturelles Engagement (vgl. [Ca09], S. 29). Der Konsumstil dieses Segments kennzeichnet sich dabei dadurch, dass Personen dieser Gruppe tendenziell eher nicht konsumorientiert sind und einen hohen Wert auf Umweltschutz legen (vgl. [Ca09], S. 29). Der Aspekt der geringen Konsumorientierung korreliert dabei mit einem geringen Wissen über die Angebote und Möglichkeiten, die der Markt dem Konsumenten bietet, da Konsum für diesen keinen hohen Stellenwert einnimmt. Genau dieser Tatbestand ist es dann, der negativ auf den effizienten Energiekonsum im Haushalt rekurriert (vgl. [Ca09], S. 29 nach [PW14], S. 28). Diese Konsumentengruppe legt Wert auf einen umweltschonenden effizienten Energiekonsum, wird jedoch nicht auf die vorliegenden Möglichkeiten aufmerksam aufgrund des gering ausgeprägten Konsuminteresses. Das Wissen um diese gesammelten Faktoren stellt dann im Weiteren ein enormes Potenzial dar, wenn es um die gezielte Ansprache dieses Kundensegments geht. So kann hier mithilfe des Wissens der aufgezeigten Determinanten, eine auf den Konsumenten zugeschnittene Ansprache erfolgen, welche dann in dessen Konsumverhalten interveniert.

Auf ähnliche Art und Weise hat auch [Pr10] in seiner Studie anhand von

soziodemografischen Faktoren und Lebensstilaspekten Konsumententypen herausgearbeitet, welche aufgeschlossen sind gegenüber der Technologie von Smart Metern und welche somit als ideale Zielgruppe für Produkte dieses Segments fungieren (vgl. [Pr10], S. 52). Somit stellt die Identifikation und Zuordnung von Konsumenten in solche Typologien einen wichtigen Aspekt für marketingtechnische Zwecke dar.

Um die Ausprägung dieser verschiedenen Elemente zu erfahren, ist es nötig, diese zunächst durch Kundenumfragen empirisch zu ergründen. Die eher latent verborgenen Werte der Konsumenten müssen dabei durch eine entsprechende Operationalisierung messbar gemacht werden. Im darauf folgenden Schritt sollen die erhobenen Daten mit den Lastgangprofilen abgeglichen werden, um auf die Weise prototypische Konsumententypologien zu erstellen, die die Grundlage für die künftige datengetriebene Identifikation und Zuordnung der verschiedenen Konsumenten bildet.

7 Fazit

Die konsumentenspezifischen Aspekte, die die verschiedenen Energiekulturen in den Konsumentenkreisen determinieren, können effektiv umgesetzt werden, um zum Beispiel zielgruppenspezifische Strategien zu entwickeln, beziehungsweise um Impulse für eine energiebezogene Verhaltensänderung zu setzen (vgl. [Sp10], S. 6). Ausgangspunkt für diese Überlegung bildet der Status, den das Produkt Energie im Konsumbereich einnimmt. Energie stellt so für den Konsumenten ein unsichtbares, im Alltag eher weniger thematisiertes Produkt dar (vgl. [Sp10], S. 6 nach [DT05], S. 6). Dementsprechend ist das Wissen und das Interesse bezüglich des Energiekonsums beim Konsumenten eher gering im Vergleich zu anderen, emotional stärker aufgeladenen Produkten. Aus diesem Grund ist eine Thematisierung des Energiekonsums effektiver, wenn diese lebensstilspezifische Vorstellungen integriert. Wird beispielsweise eine energiespezifische Nachhaltigkeitsstrategie durch die Medien kommuniziert, so trifft diese Botschaft auf jeweils konsumentenspezifische Deutungs- und Handlungskontexte, welche durch die soziale Lage des Konsumenten (Milieu, Alter, Geschlecht, Wohnsituation usw.) determiniert sind.

Um hierbei die Effizienz der Ansprache zu erhöhen, ist es von Bedeutung, die jeweilige Zielgruppe in allen Dimensionen möglichst genau zu identifizieren und anhand dieser Faktoren eine bezüglich der Zielgruppe möglichst komplementäre Strategie zu entwickeln (vgl. [Sp10], S. 6 nach [Br00]; [EGS00]; [Se00). Dieses komplementäre Verhältnis von Produktkommunikation und Konsument steigert die Präsenz des Produkts und akzentuiert zielgruppenspezifische Vorteile, welche als Anreize zum Befolgen der Botschaft fungieren. Sind beispielsweise Feedbacksysteme nicht ausreichend auf die soziale und kulturelle Vielfältigkeit bezüglich der Lebensstile der Konsumenten angepasst, kann dies das Potenzial der Energieersparnis, welche solche Systeme verfolgen, aufheben. Aus diesem Grund liefert eine interdisziplinäre Verschränkung der Wirtschaftsinformatik mit der soziologischen Perspektive wichtige Erkenntnisse über die

zielgruppenspezifischen Handlungsansätze, welche bei der Konstruktion solcher Systeme nutzbar gemacht werden können (vgl. [Sp10], S. 2, 11).

Dieses Wissen um die Bedeutung der Komplementarität von Zielgruppe und Produktkommunikation findet folgend Verwendung in Informationssystemen zur Auswertung von Energiedaten. Erste Ansätze wurden in der Literatur identifiziert und prototypisch umgesetzt. In weiteren Forschungen sollen nun ausgewählte konsumentenspezifische Merkmale in den Segmentierungsprozess einfließen. Diese werden anhand einer Lernstichprobe bei einem repräsentativen Anteil des Gesamtkundenstamms erhoben. Die Segmentierung der Lastgänge der Kunden hat dann nicht mehr ausschließlich den Fokus auf die maximale Ähnlichkeit der Lastgänge, sondern gestaltet die entstehenden Segmente auch bzgl. kundenspezifischer Merkmale intern möglichst homogen. Neu hinzukommende Kunden, von denen nur ein Lastgang bekannt ist, können dann über Klassifikationsalgorithmen den Kundengruppen zugeordnet werden, inklusive der kundenspezifischen Merkmale. Diese reichern den bislang „anonymen“ Lastgang an, und bieten der Marketingabteilung wertvolle neue Einblicke.

Ein Thema, welches in der vorliegenden Untersuchung bislang nicht referenziert wurde, aber nicht minder wichtig ist, ist der Datenschutz. Dieser wurde bislang bewusst ausgeblendet, da die Regelungen von Land zu Land unterschiedlich sind. Generell gilt in Deutschland, dass kundenbezogene Merkmale nicht ohne das Einverständnis des Kunden erhoben oder weiter verarbeitet werden dürfen. Untersuchungen zu alternativen Möglichkeiten der Verarbeitung von Daten, beispielsweise auf anonymisierter oder pseudonymisierter Grundlage, sind bereits in Planung. Insbesondere bei Bestandskunden ist dies ein wichtiger Punkt, da diese Verträge in der Regel keine Freigabe zur Datenverarbeitung enthalten, so haben Gespräche mit Energieversorgern ergeben.

8 Literatur

[AR13] Albert, A.; Rajagopal, R.: Smart Meter Driven Segmentation: What Your Consumption Says About You. IEEE Transactions on Power Systems, 28 (4), 4109-4030. 2013.

[BGS12] Bohunovsky, Lisa; Grüneberger, Sigrid; Schwarz, Julian: Energieverbruach engagierter Haushalte. Bericht zu Arbeitspaket 2 aus dem Projekt „BENE. BürgerEngagement für Nachhaltige Energie“. 3. Auflage. Österreichische Forschungsförderungsgesellschaft mbH (FFG). 2012.

[BJ05a] Bundesministerium der Justiz: Gesetz über die Elektrizitäts- und Gasversorgung (Energiewirtschaftsgesetz - EnWG). 2005.

[BJ05b] Bundesministerium der Justiz: Verordnung über den Zugang zu Elektrizitätsversorgungsnetzen (Stromnetzzugangsverordnung - StromNZV). 2005.

[Br00] Brand, Karl-Werner: „Ich lasse mir meinen Schweinsbraten nicht vermiesen“. Alltagsbewußtsein, Lebensstile und nachhaltiger Konsum, in Günther Claudia, Fischer

Corinna, Lerm Susanne (Hg.): Neue Wege zu nachhaltigem Konsumverhalten. Berlin. 2000.

[BSI14] Bundesamt für Sicherheit in der Informationstechnik: Schutzprofil für ein Smart Meter Gateway (BSI-CC-PP-0073). 2014.

[BWE15] Bundesministerium für Wirtschaft und Energie: Baustein für die Energiewende: 7 Eckpunkte für das „Verordnungspaket Intelligente Netze“. 2015.

[Ca09] Carrera, Diana et al.: Lebensstilkonzepte: Konsumanalyse – Nutzung der Wärmeenergie. Internationales Zentrum für Kultur- und Technikforschung: Universität Stuttgart. 2009.

[Ch09] Christanell, Anja: Nachhaltiger Konsum und gesellschaftliche Lebensrealitäten Ansätze zur Förderung nachhaltiger Konsummuster unter besonderer Berücksichtigung soziokultureller Einflüsse, gesellschaftlicher Trends und Kommunikationsstrategien. ÖIN – Österreichisches Institut für Nachhaltige Entwicklung: Wien. 2009.

[Ch00] Chapman, P.; Clinton, J.; Kerber; R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R.: CRISP-DM 1.0 Step-by-step data mining guides. 2000.

[Ch12] Chicco, G.: Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy 42 (2012), 68-80. 2012.

[DT05] Dobbyn, Judith; Thomas; G.: Seeing the light: the impact of microgeneration on the way we use energy. Qualitative research findings. Hub Research Consultants London. 2005.

[EGS00] Empacher, Claudia; Götz, Konrad; Schultz, Imgard: Demonstrationsvorhaben zur Fundierung und Evaluierung nachhaltiger Konsummuster und Verhaltensstile. Frankfurt. 2000.

[EY13] Ernst & Young: Kosten-Nutzen-Analyse für einen flächendeckenden Einsatz intelligenter Zähler. Hg. v. Ernst & Young. 2013.

[EY15] Ernst & Young: Stadtwerkestudie 2014 Nachhaltige Geschäftsmodelle für Stadtwerke. 2015.

[Fe06] Fettke, P.: Eine Untersuchung der Forschungsmethode "Review" innerhalb der Wirtschaftsinformatik. In: Wirtschaftsinformatik (4), S. 257–266. 2006.

[Fl12] Flath, C.; Nicolay, D.; Conte, T.; Dinther, C.; Filipova-Neumann, L.: Clusteranalyse von Smart-Meter-Daten, in: Wirtschaftsinformatik, Bd. 54, Nr. 1, S. 33–42. 2012.

[FRV05] Figueiredo, V.; Rodrigues, F.; Vale, Z.: An electric energy consumer characterization framework based on data mining techniques. IEEE Transactions on Power Systems, Vol. 20, No. 2, 596 – 602. 2005.

[GB11] Gröger, Maria; Bruckner, Thomas: Lebensstile und Investitionsverhalten im Wärmemarkt. Internationale Energiewirtschaftstagung an der TU Wien. IEWT 2011. 2011.

[HWH14] Heinrich, K.; Weiß, T., Hilbert, A.: Big Data Analytics in the Energy Market – Mining Customer Profiles from Smart Meter Data. Vortrag auf der SAS Conference Analytics

2014, 05.06.2014.

[Mo08] Moss, S. J.: Market segmentation and energy efficiency program design. CIEE Behavior and Energy Program, 2008.

[NGB07] Nipkow, Jürg; Gasser, Stefan; Bush, Eric: Der typische Haushalt- Stromverbrauch. Energieverbrauch von Haushalten in Ein- und Mehrfamilienhäusern. In: Bulletin SEV/VSE 19/2007, S. 24-26. 2007.

[Pr10] Profijt, Markus: Smart Meter – Welcher Haushalt will ihn haben? In: e | m | w. Heft 6/ 10. S. 52 – 54, 2010.

[PW91] Prose, Friedmann; Wortmann, Klaus: Die sieben Kieler Haushaltstypen – Werte, Lebensstile und Konsumverhaltensweisen. Endbericht Band 1. Kiel. 1991.

[Re99] Reusswig, Fritz: Umweltgerechtes Handeln in verschiedenen Lebensstil-Kontexten, in Linneweber, Volker, Kals, Elisabeth (Hg.): Umweltgerechtes Handeln. Barrieren und Brücken. Berlin et al. 1999.

[Se00] Sehrer, Walter: Zielgruppen und Konsum – Stile eines nachhaltigen Konsums. In: Günther, Claudia; Fischer, Corinna; Lerm, Susanne (Hg.): Neue Wege zu nachhaltigem Konsumverhalten. Berlin. 2000.

[Se12] Servatius, H.-G.; Schneidewind, U.; Rohlfing, D. (Hg.): Smart Energy. Wandel zu einem nachhaltigen Energiesystem. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. 2012.

[SEAI12] Electric Ireland and Sustainable Energy Authority of Ireland , Full Data from National Smart Meter Trial Published, http://www.seai.ie/News_Events/Press_Releases/2012/ Full_Data_from_National_Smart_Meter_Trial_Published.html, 2012.

[Sp10] Spitzer, Markus: Die Beeinflussung des Energiekonsums durch Energieverbrauchsrückmeldesysteme. Ein Appell zur Stärkung der lebensstilspezifischen Energieverbrauchsforschung. ÖIN – Österreichisches Institut für Nachhaltige Entwicklung: Wien. 2010.

[VDE15] Verband der Elektrizitätswirtschaft VDEW: VDEW Lastprofile Online abrufbar unter: http://www.gipsprojekt.de/featureGips/sw_unna/EnwgTool/Stromnetz/Standard lastprofile/VDEW-Lastprofile/VDEW-Lastprofile/Lastprofil_H0.xls. Letzter Abruf 25.04.2015.

[Wa14] Wang, P.: Anwendung von Methoden des Data Minings zur Mustererkennung aus dem Smart Metering. Masterarbeit. Technische Universität Dresden. 2014.

[Wi14] Wikipedia: Lastprofil VDEW H0 Winter, Online abrufbar unter: http://commons.wikimedia.org/wiki/File:Lastprofil_VDEW_H0_Winter.png. Letzter Abruf 25.04.2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Modellbasierte Optimierung von Energiesystemen

David Bauer[1], Stefan Kirschbaum[2], Gregor Wrobel[3], Julian Agudelo[4] und Philip Voll[5]

Abstract: Es wird ein Softwareframework vorgestellt, welches eine modellbasierte Simulation und Optimierung von Energieversorgungssystemen ermöglicht. Ziel der Entwicklung war es, komplexe mathematische und softwaretechnische Methoden Anwendern zugänglich zu machen, die zwar über profunde Kenntnis der inhaltlichen Fragestellungen verfügen, sich aber nicht mit den in der Software verwendeten Methoden auseinandersetzen können. Zu diesem Zweck wurde ein flexibles Softwareframework geschaffen, das die Anbindung verschiedener fachlicher Funktionen, wie Simulation, Optimierung oder Wirtschaftlichkeitsbewertung, ermöglicht und die methodische Komplexität hinter einer praxisnahen Benutzerführung und Oberfläche verbirgt. Es werden Praxisbeispiele beschrieben, die aufzeigen, dass diese Aufgabe erfolgreich gelöst werden konnte.

Keywords: Softwareframework, domänenspezifische Sprachen, Simulation, Optimierung, Energiesystemanalyse, Energieeffizienz

1 Einleitung

Die zunehmende Nutzung erneuerbarer Energien im Rahmen der Energiewende bei gleichzeitigem Ausstieg aus der Atomenergie bis 2022 bedingt einen raschen Wandel der Energielandschaft in Deutschland. Um die ehrgeizigen Klimaziele zu erreichen, ist eine Reduktion des Primärenergieverbrauchs durch gesteigerte Energieeffizienz unumgänglich. Immer häufiger entstehen verteilte, dezentrale Energiesysteme, um die Wärme- und Stromversorgung flexibel und damit wirtschaftlich und technisch optimal zu gestalten. Diese Systeme bergen große Optimierungspotenziale hinsichtlich ihrer Struktur und Dimensionierung.

Der Ausbau erneuerbarer Energien und die Förderung von Kraft-Wärme-Kopplung hat ebenfalls zu einer höheren Preisvolatilität an den Strommärkten geführt. Hieraus resultieren hohe Einsparpotenziale, die sich durch eine optimierte und marktorientierte Betriebsweise von Energiesystemen erschließen lassen. Die vielen interagierenden Parameter, die dabei eine Rolle spielen, ergeben ein komplexes System, bei dem mathematische Optimierungsalgorithmen genutzt werden müssen.

Zur Bewertung und Simulation von industriellen Energiesystemen werden oftmals

[1] GFaI, Graphische Ingenieursysteme, Volmerstr. 3, 12489 Berlin, bauer@gfai.de
[2] GFaI, Graphische Ingenieursysteme, Volmerstr. 3, 12489 Berlin, kirschbaum@gfai.de
[3] GFaI, Graphische Ingenieursysteme, Volmerstr. 3, 12489 Berlin, wrobel@gfai.de
[4] GFaI, Graphische Ingenieursysteme, Volmerstr. 3, 12489 Berlin, agudelo@gfai.de
[5] Currenta, Energy economics and energy projects, Chempark, 51368 Leverkusen, philip.voll@currenta.de

Tabellenkalkulationsprogramme eingesetzt. Diese bieten zwar eine große Flexibilität hinsichtlich der Modellierung, sind aber nicht speziell auf die Domäne der Energiesysteme zugeschnitten. Es fehlen Möglichkeiten, die Ergebnisse schematisch und transparent darzustellen. Die Modelle lassen sich meist nur von den Programmierern der Lösungen bedienen. Eine Verwendung von fachfremden Nutzern ist nahezu unmöglich. Zudem ist es schwierig, verschiedene Modelle von Energiewandlungsanlagen zu kombinieren. Zum Lösen komplexer Optimierungsprobleme werden spezialisierte Software-Tools benötigt.

Im Folgenden wird ein Simulationswerkzeug vorgestellt, welches es erlaubt, industrielle Energiesysteme in Struktur und Betrieb zu optimieren, und das gleichzeitig eine gute Bedienbarkeit bietet. Die Nutzung der Software durch mathematisch weniger geschultes Personal war bei der Entwicklung des Frameworks eine strategische Herausforderung. Anlagenfahrer können eine Betriebsoptimierung oder Energieberater eine Strukturoptimierung durchführen, ohne dass ein fachliches Verständnis der dahinter liegenden mathematischen Methoden und Modelle erforderlich ist.

2 Das Software-Framework TOP-Energy

Das Framework TOP-Energy [TE15] wurde über mehrere Jahre von der Gesellschaft zur Förderung angewandter Informatik und vom Lehrstuhl für Technische Thermodynamik der RWTH Aachen entwickelt. Der Lösungsansatz basiert auf folgenden Grundsätzen:

- *Modularer Softwareaufbau.* Die Software besteht aus verschiedenen Modulen zur Ersterfassung, Simulation, Optimierung und zum Vergleich von Energiesystemen. Durch Verwendung von domänen-spezifischen Modellierungssprachen ist der Entwicklungsaufwand für neue Module, z.B. zur Sensitivitäts- oder Pinchanalyse, gering.
- *Komponentenweise Simulationsmodelle.* Die Energiewandlungsanlagen und Stoffmodelle innerhalb der Simulationsmodule sind als Vorlagen verfasst, welche eine einheitliche Trennung zwischen Benutzereingabe, Ausgabe und dem fachlichen Modell besitzen. Durch Instanziieren der Templates (Komponenten) können die Bestandteile der Energiesystemanlagen mit Hilfe eines Flussschemaeditors zu einem Energiesystem zusammengefügt werden. Anwender können die bestehende Vorlagenbibliothek durch eigene Vorlagen erweitern und parametrierte Komponenten speichern.
- *Ein hybrider Simulations- und Optimierungsalgorithmus.* Der zur Simulation verwendete Algorithmus kann sowohl große, algebraische Gleichungssysteme als auch gemischt ganzzahlig lineare Optimierungsprobleme und einfache Zuweisungen lösen. Die Aufteilung des Problems zwischen Simulation und Optimierung ist abhängig vom Energiesystem und der Aufgabenstellung und ergibt sich aus den Fragestellungen der Anwender.

- *Auf die Energietechnik zugeschnittene Benutzerschnittstellen.* Um eine ausgezeichnete Bedienbarkeit zu gewährleisten, sind Editoren zur Bearbeitung von energietechnischen Lastgängen und Kennlinien erforderlich. Für die übersichtliche Darstellung der Simulationsergebnisse bietet das Flussschema spezielle Ansichten, mit denen Massen-, Enthalpie und Exergieströme im Energiesystem Sankey-artig visualisiert werden können.

2.1 Grundlegende Architektur

Das TOP-Energy-Framework basiert auf dem ForGE-Framework zur Entwicklung von Applikationen für graphische Ingenieursysteme. Dieses enthält Funktionalität, um netzartig und hierarchisch strukturierte Applikationsdaten aufzunehmen und zu visualisieren. Dazu gehören auch Layout-Verfahren zum automatischen Platzieren und Routen in schematischen Darstellungen. In Forschungsprojekten wurden weitere ForGE-Applikationen entwickelt, beispielsweise zur Wissensverarbeitung im Qualitätsmanagement oder zur Simulation von Produktionsprozessen.

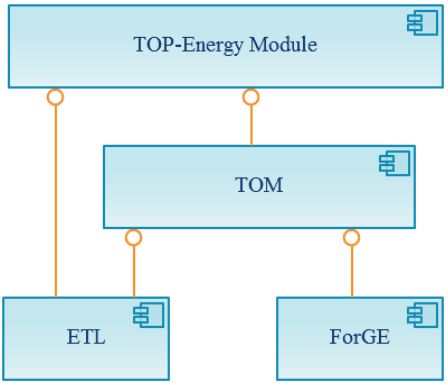


Abb. 1: Softwarekomponenten in TOP-Energy

Das ForGE-Framework wurde um die Engineering-Template-Library (ETL) ergänzt. Hierbei handelt es sich um eine Bibliothek, welche umfangreiche Berechnungen mit ingenieurtechnischen Datentypen ermöglicht. Die ETL wurde ursprünglich am Lehrstuhl für Technische Thermodynamik der RWTH Aachen entwickelt. Zu den wichtigsten Datentypen zählen Zahlen mit Einheit, Terme, Zeitreihen und Kennlinien.

Das Speichern der Daten erfolgt im Datenmodell des ForGE. Die dort enthaltene aber stark abstrakte Datenstruktur und die in der ETL implementierten Funktionen zur Berechnung werden vom TOP-Energy Object Model (TOM) überbrückt. Das TOM stellt gleichzeitig eine fachliche Schnittstelle für die Modulprogrammierung dar. Die Daten werden durch zusammengesetzte Datenobjekte implementiert, welche nach außen wie ETL-Datentypen erscheinen, intern aber als abstrakter Datenbaum in ForGE serialisiert

werden. Somit besteht eine Trennung zwischen dem internen Datenmodell und den fachlichen Modulen der Software durch das TOM.

Die Module werden dynamisch zur Laufzeit über die Kommandoschnittstelle des ForGE in das System integriert. Auf diese Art und Weise können Modulfunktionalitäten erweitert und modifiziert werden ohne in die Softwarebasis einzugreifen. Die zugrunde liegenden Mechanismen basieren auf entsprechenden Entwurfsmustern. Um die Modulerstellung zu beschleunigen, gibt es ähnlich wie für die Komponenten Modellierungssprachen und -editoren.

2.2 Modularer Aufbau

Standardmäßig sind folgende Module verfügbar:

- *eNtry.* Dieses Modul dient einer ersten energieorientierten Analyse eines Energiesystems und erfasst die wichtigsten Rahmendaten in einem Fragebogen. Daraus werden betriebliche Energiekennzahlen ermittelt und typischen Branchenwerten gegenüber gestellt.
- *eSim.* Hiermit lassen sich Energiesysteme unter Vorgabe von zeitlichen Lastgängen sowie von technischen und wirtschaftlichen Randbedingungen simulieren. Das Simulationsmodell wird in einem Flussschemaeditor zusammengestellt, welcher gleichzeitig der übersichtlichen Präsentation der Simulationsergebnisse dient.
- *eSynthesis.* Entsprechend lassen sich mit diesem Modul Energiesysteme unter Vorgabe von Randbedingungen hinsichtlich ihrer Struktur, ihrer Dimension oder ihres Betriebs optimieren.
- *eVariant.* Das Modul sammelt bewertungsrelevante ökonomische und ökologische Größen eines Simulationsmodells und illustriert Zusammenhänge in einer Übersicht.
- *eValuate.* Zum Vergleich von verschiedenen Varianten eines Energiesystems lässt sich hiermit eine Wirtschaftlichkeitsrechnung durchführen.

Die verschiedenen Module werden durch einen Zeitreiheneditor zum Bearbeiten von energietechnischen Lastgängen ergänzt. Dieser kann auch als eigenständige Applikation benutzt werden und ist an TOP-Energy angebunden. Der Zeitreiheneditor bietet verschiedene Verfahren zum Erzeugen, Bearbeiten, Filtern, Interpolieren und Aggregieren von Lastgängen.

2.3 Komponenten als Bausteine für Simulationsmodelle

Die Simulationsmodelle bestehen aus Komponenten, welche technische Anlagen, Stoffströme, Energiebedarfe, Tarife oder Außengrenzen eines Energiesystems

repräsentieren. Auf diese Weise kann man verschiedene Varianten eines Systems schnell erstellen und verändern.

Die Eingabe und Ausgabe von Lastgängen und Parametern erfolgt über Formulare und ist in allen Komponenten einheitlich von der Modellbeschreibung separiert.

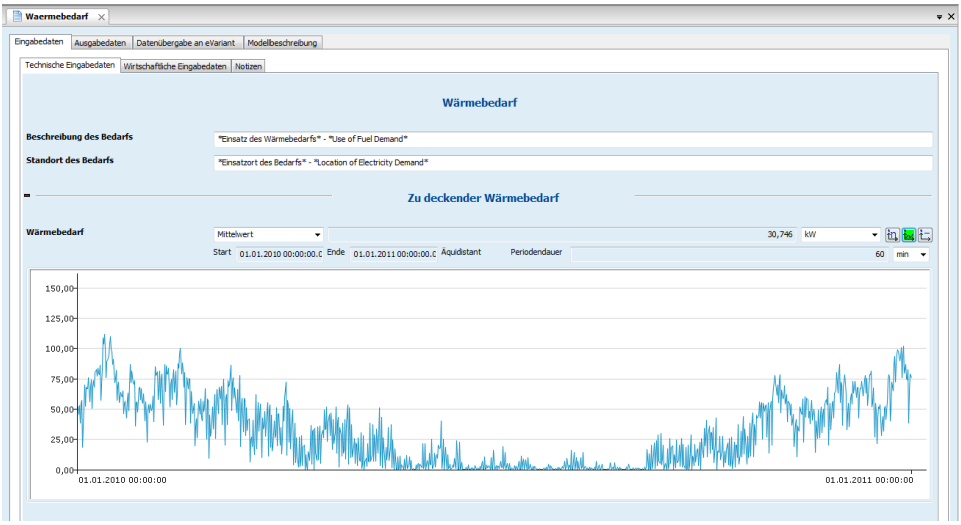


Abb. 2: Eingabeformular der Komponente „Wärmebedarf“

Die Ein- und Ausgabeformulare sind zeilenbasiert aufgebaut. Jede Zeile repräsentiert ein TOM-Datenobjekt und wird durch eine Menge von grafischen Steuerelementen realisiert. Der Aufbau der Formulare, aber auch die Sichtbarkeit von Bedienelementen in Abhängigkeit von Datenwerten, wird in einer eigenen Formularbeschreibungssprache definiert. Dabei handelt es sich um ein XML-Derivat, das zur Beschreibung von Oberflächen konzipiert wurde. Diese Beschreibung wird zur Systemlaufzeit interpretiert, und die jeweiligen Formulare generiert. Analog zur Beschreibung vom Formularen gibt es die Möglichkeit, Reports durch eine eigene Beschreibungssprache zu spezifizieren. Die Verwendung von XML-Sprachen stellte sich als eine geeignete Form zur Beschreibung von Formularen und Berichten durch versierte Anwendern dar und ermöglicht neue technische Komponenten hinzuzufügen oder bestehende Komponenten nach den eigenen Anforderungen zu modifizieren und das ohne Programmierkenntnisse.

Die mathematische Beschreibung der Komponenten erfolgt mit Hilfe einer textuellen Modellierungssprache (PML - Process Modelling Language), welche an Modelica [Mo15] angelehnt ist. Benötigte Variablen müssen zunächst mit Name und Einheit deklariert und können ggf. mit Steuerelementen der Ein- und Ausgabeformulare verknüpft werden. Als Variablentypen stehen neben kontinuierlichen, binären und ganzzahligen Variablen auch Kennlinien und Listen zur Verfügung. Die physikalischen, technischen und wirtschaftlichen Zusammenhänge werden anschließend in Form von

Gleichungen beschrieben. So enthält die Komponente BHKW relevante technische Daten wie Nennleistung oder Wirkungsgrad und ein Modell, das die Umwandlung der chemischen Energie des Brennstoffs in thermische und elektrische Energie beschreibt.

2.4 Flussschemaeditor

Im Flussschemaeditor werden die verschiedenen Komponenten zu einem Energiesystem verknüpft, indem technische Anlagen, Tarife und externe Einflüsse durch Stoffströme miteinander verbunden werden. Das Layout der Netze erfolgt durch ein entwickeltes Routingverfahren. Hierbei handelt es sich um ein orthogonales Routing, welches für Hyperkanten eine möglichst durchgängige geradlinige Struktur berechnet.

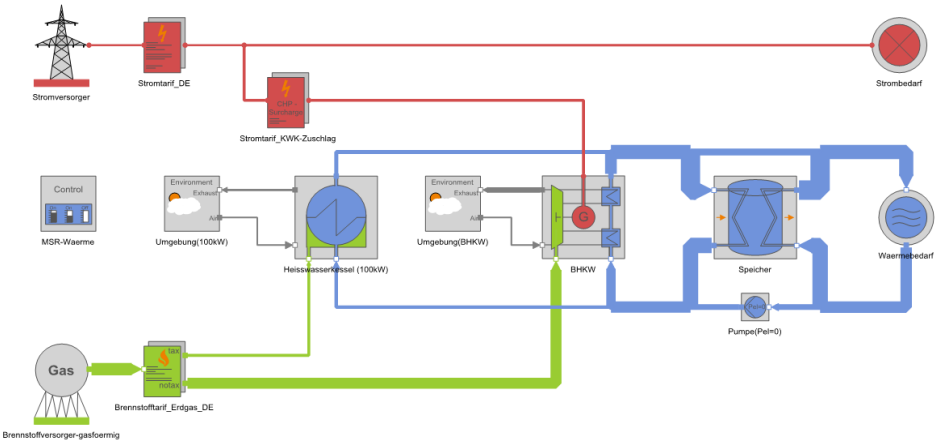


Abb. 3: Flussschema eines Energiesystems mit Busrouting

Zur Visualisierung der Ergebnisse lassen sich Massen-, Enthalpie und Exergieströme veranschaulichen, indem die Kantendicken proportional zu den jeweiligen Flussgrößen dargestellt werden.

3 Simulation und Optimierung von Energiesystemen

TOP-Energy verfügt über einen hybriden Simulations- und Optimierungsalgorithmus, der in der Lage ist, die graphisch modellierten Energiesysteme in Gleichungssysteme zu konvertieren und diese zu lösen [Au05], [Ki12]. Für jeden Zeitschritt des Betrachtungszeitraums wird ein quasistationärer Zustand des Energiesystems berechnet. Der Optimierungsanteil besteht aus einem gemischt ganzzahlig linearen Optimierungsproblem, welches in der Betriebsoptimierung für jeden Zeitschritt und in der Strukturoptimierung einmal für den gesamten Simulationszeitraum gelöst wird. Der Simulationsteil wird mit Hilfe eines modifizierten Newtonalgorithmus gelöst. Zur

Verbesserung der Konvergenz steht eine Blockzerlegung und ein symbolischer Vereinfacher zur Verfügung.

3.1 Simulationsablauf und Algorithmen

In jeder Simulation werden nacheinander die folgenden Schritte ausgeführt:

1. *Einlesen der Modellbeschreibungen.* Als Erstes werden die in PML formulierten Modellbeschreibungen geparst und in ein großes Gleichungssystem überführt.

2. *Blockzerlegung.* Das entstandene Gleichungssystem, welches sämtliche Modellgleichungen beinhaltet, wird in einzelne Blöcke zerlegt, die sukzessive gelöst werden können. Die daraus resultierenden, kleineren Gleichungssysteme sind im Idealfall alle wohlbestimmt; das heißt die Anzahl der Ergebnisvariablen stimmt in jedem Gleichungssystem mit der Anzahl der Gleichungen überein. Es können aber auch unter- oder überbestimmte Systeme entstehen, z.B. durch fehlende bzw. redundante Eingabedaten oder verbleibende Freiheitsgrade im Energiesystem. Für die Blockzerlegung wird der graphentheoretische Algorithmus von Dulmage und Mendelsohn [DM58], [DM69] verwendet, welcher die Knoten eines bipartiten Graphen partitioniert. Dies entspricht der Zerlegung in einen wohlbestimmten, unterbestimmten und überbestimmten Anteil. Die weitere Aufteilung in kleine, wohlbestimmte Blöcke geschieht mit dem Algorithmus von Tarjan [Ta97], welcher die Adjazenzmatrix des wohlbestimmten Anteils in eine untere Dreiecksform transformiert.

3. *Symbolisches Vereinfachen.* Das numerische Lösen der kleinen Gleichungssysteme für jeden Zeitschritt ist verhältnismäßig zeitaufwendig und mit Fehlern behaftet. Da die Ergebnisse eines Gleichungssystems als Eingabeparameter für das jeweils nächste Gleichungssystem dienen, akkumulieren sich numerische Ungenauigkeiten durch Fehlerfortpflanzung. Sofern die Gleichungen sich auflösen lassen, werden daher einfache Gleichungssysteme symbolisch gelöst. Die gefundene Lösung wird als Binärcode kompiliert, um die Ausführung für jeden Zeitschritt zu beschleunigen.

4. *Lösen der Gleichungssysteme.* Das eigentliche Lösen der Gleichungen macht den Hauptteil der Simulation aus. Jeder Block wird zunächst für alle Zeitschritte berechnet, bevor das nächste Gleichungssystem bearbeitet wird. Für die symbolisch vereinfachten Systeme wird der generierte Binärcode ausgeführt. In der Regel verbleiben einige wenige Blöcke, die numerisch gelöst werden. Hier wird ein Newton-Raphson-Verfahren eingesetzt, welches frühzeitig abbricht, wenn sich die Qualität der Lösung nicht hinreichend verbessert und einen Neustart mit randomisierten Startwerten ausführt. Die Wahl sinnvoller Startwerte erweist sich als besonders schwierig. Sofern vorhanden, werden Lösungen aus früheren Zeitschritten als Startwerte verwendet. Zusätzlich gibt es einen Block, der das

lineare Optimierungsproblem beinhaltet. Zum Lösen dessen sind verschiedene Solver über eine Software-Schnittstelle angebunden.

3.2 Optimierungsproblem und Löser

Das Optimierungsproblem ist als gemischt ganzzahlig lineares Problem (GGLP oder auch MILP) formuliert, da diese Problemklasse robust lösbar ist und sich umfangreiche Modelle mit kommerziellen Lösern wie GUROBI oder CPLEX in akzeptabler Rechenzeit optimieren lassen [Sc98]. Es setzt sich aus kontinuierlichen und binären Variablen zusammen, die über lineare Gleichungen und Ungleichungen in Beziehung stehen. Während das gesamte Energiebilanzmodell durch kontinuierliche ($x \in \mathbb{R}^n$) Variablen realisiert wird, dienen die Binärvariablen ($y \in \{0,1\}^m$) zur Beschreibung von Zustandsgrößen. Diese geben z.B. an, ob ein BHKW in Betrieb ist oder nicht. Das Teillastverhalten der technischen Anlagen wird über linearisierte Kennlinien abgebildet.

Die allgemeine Form eines GGLP lautet:

$$f(x,y) \rightarrow \min \text{ sodass } g(x,y) = 0 \text{ und } h(x,y) \leq 0 \text{ für } x \in \mathbb{R}^n, y \in \{0,1\}^m$$

Die zu minimierende Zielfunktion $f(x,y)$ repräsentiert die Summe aller Energiekosten und –erlöse sowie etwaige Wartungs- oder Personalkosten. Zum Lösen des Optimierungsproblems stehen drei verschiedene Solver zur Verfügung, die über eine Software-Schnittstelle angebunden sind. CBC ist ein frei verfügbarer Löser aus dem COIN-OR Framework [Co15] und liefert nur für kleinere Probleme praktikable Rechenzeiten. Eine deutlich bessere Performance erzielt SCIP, der beste akademische GGLP-Löser [Ac09]. Schließlich gibt es eine Anbindung für GUROBI, einen kommerziellen, hochperformanten Löser [Gu15].

4 Praxisbeispiele

Die Funktionalität zur Struktur- und Betriebsoptimierung soll an ausgewählten Praxisbeispielen erläutert werden. Das dafür verwendete Modul eSynthesis ist im Rahmen des Forschungsprojektes *„sOptimo: Strukturoptimierung von Energieversorgungssystemen – am Beispiel von Industrieparks“* entstanden [Vo12], [Ki13]. Das Projekt wurde im Rahmen des 5. Energieforschungsprogramms der Bundesregierung gefördert.

4.1 Zeitlich aufgelöste Betriebsoptimierung

Ein österreichischer Energieversorger verfügt über ein Fernwärmenetz, dessen Betriebsweise sehr viele Freiheitsgrade enthält. Es werden drei Gasmotoren mit einer elektrischen Leistung von 4,6 MW betrieben. Des Weiteren stehen drei

Heißwasserkessel und ein Wärmespeicher zur Verfügung, der es erlaubt, Erzeugung und Bedarf zeitlich zu entkoppeln. Zusätzlich besteht die Möglichkeit, Abwärme von einem Zementwerk zu beziehen.

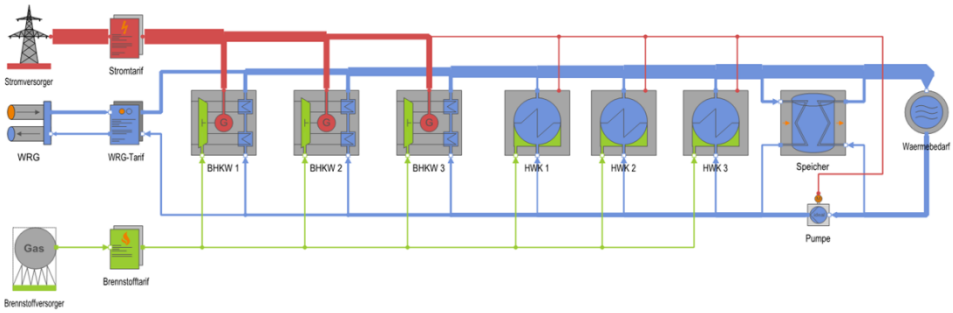


Abb. 4: Schematischer Aufbau des Fernwärmesystems

Die Auswahl der wirtschaftlichsten Fahrweise stellt eine große Herausforderung dar. Bisher wurde der tägliche Einsatzplan vom Kraftwerksmeister, basierend auf dessen Erfahrungswissen und einigen Eckdaten, am Vortag festgelegt. Aufgrund der Heterogenität des Systems und der sich schnell verändernden Marktsituation war zu erwarten, dass hier der Einsatz eines Simulationswerkzeugs zur Optimierung des täglichen Fahrplans lohnenswert ist.

Für den beschriebenen Anwendungsfall soll eine Betriebsoptimierung unter wirtschaftlichen Gesichtspunkten entwickelt werden. Das Modell sollte einerseits komplex genug sein um alle maßgeblichen Eigenschaften des Energiesystems abzubilden, andererseits aber auch benutzerfreundlich und schnell, sodass es vom Betriebspersonal in den täglichen Arbeitsprozess integriert werden kann. Als Ergebnis soll der Anwender einen Einsatzplan der verschiedenen Wärmeerzeugungseinheiten für den folgenden Tag erhalten.

Im Optimierungsmodell werden die folgenden Betriebsparameter und Faktoren berücksichtigt:

- *Wirtschaftliche Faktoren:* Strom- und Gaspreise, Betriebskosten des BHKW, Instandhaltungs- und Wartungskosten, CO_2–Kosten, KWK-Bonus, Preise für die Wärmerückgewinnung
- *Technische Parameter:* Tagesganglinie für den Wärmebedarf in Abhängigkeit von der Außentemperatur und dem Wochentag, Wirkungsgrade und Teillastverhalten der Anlagen, elektrischer Eigenverbrauch, maximale Starts pro Tag und Mindestlaufzeit der Gasmotoren, Ladezustand des Speichers

Nicht enthalten sind Investitionskosten, Kosten für das Betriebspersonal sowie sonstige Fixkosten.

Im Modell wird eine zeitliche Auflösung des Tages in 24 Stunden verwendet. Die zu minimierende Zielfunktion $f(x,y)$ repräsentiert die Summe aller Kosten und Erlöse, die bei der Bereitstellung der Wärme entstehen. Für jede Stunde des Tages setzen sich diese wie folgt zusammen:

$$C_s(x,y) + Z_B(x,y) + Z_W(x,y)$$

Dabei bezeichnet C_S die Energiekosten (für Brennstoff, Strom und Wärme aus dem Zementwerk sowie Einnahmen für die Stromeinspeisung), Z_B die Betriebskosten (wie Schmierstoffe und Harnstoff) und Z_W die Wartungskosten (aus dem Wartungsvertrag mit dem Hersteller). Die Zielfunktion entspricht der Summe der Kosten aller 24 Stunden:

$$f(x,y) = \sum_{t=1}^{24} C_S(x_t,y_t) + \sum_{t=1}^{24} Z_B(x_t,y_t) + \sum_{t=1}^{24} Z_W(x_t,y_t) \rightarrow \min$$

$$\text{sodass} \quad g(x_t,y_t) = 0 \quad \text{und} \quad h(x_t,y_t) \leq 0 \quad \forall t \in \{1,\dots,24\}$$

Die Nebenbedingungen beinhalten Energiebilanzen und Betriebsgrenzen, die eingehalten werden müssen. Sie werden für jeden Zeitschritt des zu optimierenden Tages in das Modell aufgenommen. Einige Nebenbedingungen in dem Modell koppeln die Variablen unterschiedlicher Zeitschritte miteinander. Das betrifft zum einen den Füllstand des Wärmespeichers, der zum Anfang des einen und zum Ende des nächsten Zeitschritts gleich sein muss, und zum anderen die Restriktionen für die Starts der Motoren (maximal 2 pro Tag), die alle Zeitschritte miteinander koppeln:

$$\tilde{g}(x,y) = 0 \quad \text{und} \quad \tilde{h}(x,y) \leq 0 \quad \text{mit} \quad x = (x_1,\dots,x_{24}), \quad y = (y_1,\dots,y_{24})$$

Im Optimierungslauf werden 24 Zeitschritte mit einer Schrittweite von 60 Minuten simultan berechnet. Durch die Optimierung aller Zeitschritte eines Tages in einem einzigen Gleichungssystem können zeitlich koppelnde Randbedingungen berücksichtigt werden. Die Einbeziehung einer vorausschauenden Speicherbe- und -entladung und einer Reduktion der Starts durch eine prädiktive Fahrweise der Motoren macht zwar das Optimierungsproblem sehr komplex, kann aber die Energie- und Kosteneffizienz deutlich verbessern. Das endgültige Problem, welches gelöst wird, beinhaltet über 5000 Variablen und ca. 8300 Gleichungen bzw. Ungleichungen.

Zur Vereinfachung der Handhabung wurde eine Schnittstellenkomponente generiert, welche die täglich zu ändernden Ein- und Ausgabedaten auf einem Formular vereint. Dies sind der Wochentag, die mittlere Außentemperatur, der stündlich aufgelöste Strompreis an der Strombörse sowie die Verfügbarkeiten der einzelnen Anlagen an dem entsprechenden Tag.

Die Ergebnisse wurden anhand ausgewählter Betriebsdaten überprüft. Die größte Herausforderung war die richtige Einbindung des Speichers, da alleine die Festlegung der Randbedingungen zu Beginn und zum Ende des Betrachtungszeitraumes eine betriebliche Herausforderung darstellt. Das Modell läuft seit einigen Monaten absolut

stabil. Die Simulation liefert eine sehr hohe Übereinstimmung des prognostizierten mit dem tatsächlichen Wärmebedarf. Der Fehler des Tagesmittelwerts liegt für gewöhnlich unter 2 %. Ein vorsichtiger erster Vergleich der bisherigen, händischen Fahrweise mit den Einsatzplänen des Optimierungsprogrammes ergibt Kosteneinsparungen von über 10%.

4.2 Strukturoptimierung eines Industriestandorts

Ein historisch gewachsener Standort eines Pharmaunternehmens soll im Zuge der Einführung eines neuen Produktionsprozesses optimiert werden. Der Komplex besteht aus einem Haupt- (A) und einem Nebenstandort (B), die beide durch eine öffentliche Straße separiert sind. Auf Standort A sind alle Gebäude über ein zentrales Wärme- und Kältenetz miteinander verbunden. Standort B besitzt derzeit keinen Bedarf an Kälteenergie und ist daher nur mit dem Wärmenetz verbunden. Zum Betrieb des neuen Produktionsprozesses werden hier künftig 7 GWh Kälte jährlich benötigt. Insgesamt ergeben sich damit jährliche Energiebedarfe von 48 GWh für Strom, 28 GWh für Wärme und 27 GWh für Kälte.

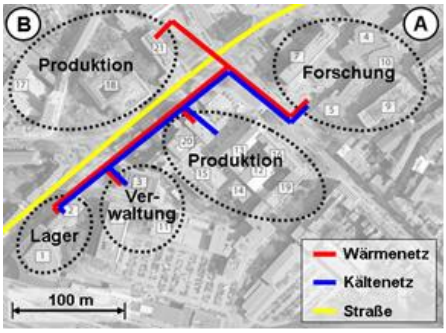


Abb. 5: Derzeitiger Gebäudeplan des Industriestandorts

Im bestehenden Komplex sind drei Kessel, ein Motorheizkraftwerk (MHKW) und drei Kompressionskältemaschinen (KKM) installiert. Als Alternativen stehen neue Kessel, MHKWs, elektrisch betriebene KKM und thermisch betriebene Adsorptionskältemaschinen (AKM) zur Verfügung. Wegen Baubeschränkungen durch die öffentliche Straße ist die nachträgliche Ankopplung von Standort B an das zentrale Kältenetz nicht möglich. Die verschiedenen Entscheidungsmöglichkeiten für oder gegen neue Anlagen werden durch zusätzliche Binärvariablen im Optimierungsmodell repräsentiert.

Ausgehend vom Ist-Fall wird die wirtschaftlich optimale Versorgungslösung gesucht, die den Kapitalwert

$$C = Z \cdot \frac{(i+1)^T - 1}{(i+1)^T \cdot i} - I$$

maximiert. Hierbei bezeichnet Z den Zahlungsstrom pro Periode und I die gesamten Investitionskosten. Es wird ein Zeitraum T von 10 Jahren mit einem internen Zinssatz i von 8 % betrachtet. Der Kapitalwert ist in der Regel negativ, da die Energiebezugskosten die Erlöse durch Stromeinspeisung übersteigen.

Die beste Lösung stellt einen Kompromiss zwischen Investitions- und betriebsgebundenen Kosten her und ermöglicht jährliche Energie- und Wartungskostenersparnisse in Höhe von 4,73 Mio €. Der Kapitalwert ist um 39 % besser als im Ist-Fall. Es werden lediglich ein Kessel und eine Kältemaschine aus dem aktuellen Bestand behalten. Alle übrigen Anlagen werden durch neue effizientere Anlagen ersetzt.

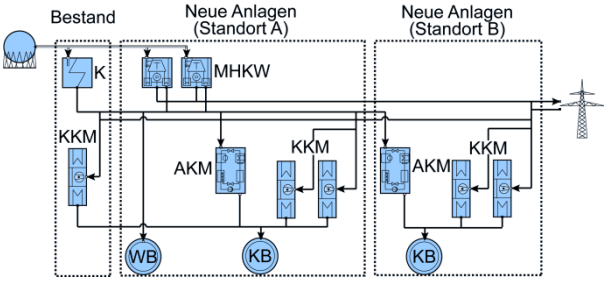


Abb. 6: Optimierte Anlagenstruktur

Die Anlagenverschaltung und –dimensionierung sind so gewählt, dass die neu installierten Anlagen fast vollständig in Teillastbereichen mit maximaler Effizienz betrieben werden.

Eine genauere Analyse des mit der Umrüstung des Standorts beauftragten Ingenieurbüros ergab, dass aus platztechnischen Gründen keine Lösung mit mehr als zwei MHKWs realisiert werden kann. Das Büro erarbeitete anhand des Vorschlags einen dreigliedrigen Umrüstungsplan, der zunächst die Anlagen zur Versorgung von Standort B installiert.

5 Fazit und Ausblick

Die vorgestellte Softwareplattform zur Simulation und Optimierung von Energiesystemen ermöglicht es Anwendern, komplexe Fragestellungen für Energiesysteme zu bearbeiten. Anhand von zwei Praxisbeispielen wurde nachgewiesen, dass das Framework durch Kapselung der dahinter liegenden mathematischen Modelle auch fachfremden Nutzern zugänglich ist. Um künftig mehr Flexibilität beim Erstellen der Modelle zu erreichen, soll die Bibliothek der verfügbaren Energiesystem-komponenten sukzessive erweitert werden. Für die Synthese wären beispielsweise Komponenten für Solarthermie und Dampferzeugung wichtig.

Literaturverzeichnis

[Ac09] Achterberg, T.: SCIP: solving constraint integer programs. Mathematical Programming Computation Vol. 1/1, S. 1-41, 2009.

[Au05] Augenstein, E.; Herbergs, S.; Kuperjans, I.; Lucas, K.: Simulation of industrial energy supply systems with integrated cost optimization. In (Kjelstrup, S.; Justad, J. E., Hrsg.): Proceedings of ECOS 2005: 18th International Conference on Efficiency, Cost, Optimization, Simulation, and Environmental Impact of Energy Systems. Trondheim, Norway, S. 627-634, 2005.

[Co15] COIN-OR, www.coin-or.org, Stand: 30.04.2015.

[DM58] Dulmage, A. L.; Mendelsohn, N. S.: Coverings of bipartite graphs. Canad. J. Math 10, S. 517-534, 1958.

[DM69] Dulmage, A. L.; Mendelsohn, N. S.: Some Graphical Properties of Matrices with Non-Negative Entries. Aequationes mathematicae 2, S. 150-162, 1969.

[EP15] European Power Exchange Marktdaten, www.epexspot.com/de/marktdaten/, Stand: 30.04.2015.

[Gu15] Gurobi, www.gurobi.com/de, Stand: 30.04.2015.

[Ki12] Kirschbaum, S.: Entwicklung eines Softwarepaketes zur Simulation industrieller Produktionsprozesse unter energetischen Gesichtspunkten. Shaker Verlag, 2012.

[Ki13] Kirschbaum, S.; Voll, P.; Scheffler, R.; Plessow, M.; Bardow, A.: Simultaneous optimization of structure and infrastructure for distributed energy supply systems. In (Xu, J.; Jin, H.; Zhan, N., Hrsg.): Proceedings of ECOS 2013: 26th International Conference on Efficiency, Cost, Optimization, Simulation, and Environmental Impact of Energy Systems. Guilin, China, S. 1-10, 2013.

[Mo15] Modelica, www.modelica.org, Stand: 30.04.2015.

[Sc98] Schrijver, A.: Theory of Linear and Integer Programming. John Wiley & Sons, 1998.

[Ta97] Tarjan, R.: Depth first search and linear graph algorithms. SIAM Journal on Computing 1/2, S. 146-160, 1997.

[TE15] TOP-Energy, www.top-energy.de, Stand: 30.04.2015.

[Vo12] Voll, P.; Lampe, M.; Wrobel, G.; Bardow, A.: Superstructure-free synthesis and optimization of distributed industrial energy supply systems. Energy 45/1, S. 424-435, 2012.

[WKP12] Wrobel, G.; Kirschbaum, S.; Pleßow, M.: Modellgetriebene Entwicklung von Softwarelösungen für die Energietechnik. Praxisforum Modellierung 2012, Otto-Friedrich-Universität Bamberg, 2012.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

A Fully Distributed Continuous Planning Approach for Decentralized Energy Units

Astrid Nieße[1] and Michael Sonnenschein[2]

Abstract: In this contribution, we present the development of DynaSCOPE as a fully distributed continuous planning approach for decentralized energy units organized in Dynamic Virtual Power Plants (DVPP). The work at hand elaborates on possible basic algorithms for the DVPP planning task, motivates the choice of a fully distributed algorithm and gives details and evaluation results for the presented system. It is shown, that DynaSCOPE meets the requirements for the DVPP planning task and effectively enhances product delivery in case of energy unit outages or prognosis deviations.

Keywords: distributed algorithms, Smart Grid, virtual power plants, DynaSCOPE, COHDA

1 Introduction

Distributed algorithms for the control of real-world components are often considered for application motivated by one of the following aspects: Either the complexity of the control tasks shows an abrupt rise or new requirements are given regarding functionality or performance. In both cases, a conventional extension of the control system at hand might not be applicable. In the context of the electrical energy system both aspects are true: With the still rising amount of distributed energy ressources (DER), controllable loads and electrical storage devices the complexity of the control task escalates. The flexibility retrieved with these components is needed for an active operation of the system itself as a new functional requirement. Consequently, a large body of research has emerged from the field of distributed algorithms and distributed artificial intelligence in this domain within the last years.

One of the concepts presented is the Dynamic Virtual Power Plant (DVPP) [Ni12]. DVPPs are set up as product specific aggregations of distributed energy units and thus reflect the different products that can be handled at the energy markets. DVPPs can be understood as an extension to the static VPP concept presented about 15 years ago. Within DVPPs, each energy unit is represented by a software agent. These agents form DVPPs and control their energy product delivery during operation. Products are defined as an energy amount fed into the grid within a defined time span. Operation schedules are discretized by so-called planning intervals covering 15 minutes. Functional requirements have been deduced from this task: As the energy units are prone to prognosis deviations and outages (i.e. units do not follow the defined operation schedules), a continuous planning approach is needed to

[1] OFFIS - Institute for Information Technology, Escherweg 2, D-26121 Oldenburg, Germany, niesse@offis.de
[2] University of Oldenburg, D-26111 Oldenburg, Germany, michael.sonnenschein@uni-oldenburg.de

ensure product delivery in case of these incidents. This approach should detect incidents and trigger a replanning, if the product cannot be delivered as contracted at the market. Costs for the operation of the unit have to be reflected in the planning process, as well as soft constraints of the energy unit, like avoiding frequent switching of some DER. As the DVPP concept follows a self-organizing and distributed approach, a central concept would sacrifice this paradigm. In the work presented here, DynaSCOPE as a decentral approach for replanning of energy units in DVPPs is proposed. The underlying optimization problem is solved as a distributed constraint optimization problem. DynaSCOPE has been developed following the iterative process model Smart Grid Algorithm Engineering (SGAE) [NTS13] proposed for the development of (distributed) algorithms in the Smart Grid application domain. In terms of SGAE, the results of a first iteration are presented.

The rest of this contribution is structured as follows: In section 2 we will introduce some known distributed algorithms and discuss, to what extend they might be used for the task of continuous energy planning. We will then choose an appropriate algorithm and elaborate on the extension needs. Details on implementing the product specific DVPP approach in the planning process are given in section 4, followed by a description of the incident detection and replanning process in section 5. Finally we will present some evaluation results and conclude with remarks on future work.

2 Agent-based control concepts for Smart Grid control issues

The work presented by Akkermans, Ygge and Gustavsson on 1996 has been one of the first applications of agent-based control in the electrical energy system [AYG96]. The so-called HomeBots approach revealed the expected capabilities regarding scalability, flexibility, adaptivity and broad applicability [Gu99]. Since this work, many agent-based approaches habe been developed in the disciplines of electrical engineering, control and system theory and information technology and information systems. Some projects focus on algorithms like the work presented by Rahwan [RRJ09], others on compatibility with the current energy economy and roles or automation systems [ZV12]. The understanding of what constitutes a distributed system differs a lot as well: Some projects use software agents only as a concept to realize a energy unit gateway, some realize hierarchical systems [Le10] and others present fully distributed algorithms [HLS14]. In the following, we will focus on an selection of algorithms that might be used as a basic algorithm for the continuous energy planning within DVPPs.

2.1 Choosing the Right Basic Algorithm

In table 1 a selection of distributed concepts is presented. All of them have already been evaluated for a Smart Grid control task. For each concept it is depicted, if the approach already fulfils the requirement listed on the left side of the table (✓), if an extension would be needed (E) or if no extension is feasible to fulfil the specific requirement (×). The Holonic Virtual Power Plant (Hol. VPP) presented in [Tr10] has already been implemented for the

continuous planning process in static VPPs. Although products have not been reflected, an extension would be possible. The agents presented in the concept though are not capable of evaluating the quality of a new VPP schedule – this task is performed by a dedicated agent. The same aspect holds for the Autonomous Virtual Power Plants [An10]. Thus, both concepts are not applicable for the task at hand following the distributed DVPP paradigm. ALMA [Po13] is a fully distributed and highly dynamic approach implemented for a dynamic supply-demand-matching task. The introduction of products though is not possible. With COHDA [HLS14] a decentral heuristic has been proposed lately. As can be seen in table 1 all requirements for continuous energy planning can be met by extending the algorithm. In the following, we will present COHDA and discuss the extension needs for DVPP operation.

Requirement	**Hol. VPP**	**AVPP**	**ALMA**	**COHDA**
Products	E	E	×	E
Incidents	✓	(✓)	(✓)	E
Continuous planning	✓	(✓)	(✓)	E
Costs	✓	✓	✓	E
Soft constraints	✓	E	E	E
Decentral evaluation	×	×	(✓)	✓

Tab. 1: Requirements for continuous energy planning and distributed approaches. ✓: approach fulfils requirement. E: approach does not fulfil requirement, extension feasible. ×: extension not feasible.

2.2 COHDA

COHDA is a cooperative heuristic distributed at the algorithm level [HLS14]: Agents exchange information regarding their independently working algorithms to determine the optimal solution of a defined problem. The problem is to minimize the differences between an energy amount delivered by a set of energy units and a given target schedule. We follow [Ta09] in presenting the characteristics of COHDA regarding this type of algorithm. In the work presented here, COHDA should be adapted to solve a distributed constraint optimization problem. Therefore, we follow [Ch11] to describe relevant characteristics regarding this kind of problem. Implemented with agents as means for distribution, we give a short description on the agent model as well. A formal description of COHDA is given in [HLS14]; for a detailled description of COHDA see [Hi14] (German only).

- **Information exchanged:** COHDA has been developed for the day-ahead planning of clusters of energy units using agents. The agents send their working memory κ, consisting of the target schedule ζ, the known system state and the current solution candidate γ. With the target schedule ζ, each energy unit is provided an operation schedule. The known system state is used to store the current knowledge regarding the chosen operation schedules of all agents in the system.

- **Exchange criterion:** Agents exchange their working memory κ if any changes have been performed to this memory.
- **Exchange topology:** Agents communicate using a virtual overlay network, i.e. a virtual communication topology. Each agent communicates with a subset of all agents called neighborhood.
- **Information integration:** When an agent receives a message from another agent holding the working memory of this agent, he integrates all information regarding chosen operation schedules and current solution candidate γ.
- **State evaluation:** Each agent is able to evaluate a solution candidate, i.e. a cluster schedule, with respect to the target schedule ζ. All agents within the system will yield the same value for the same solution candidate.
- **Decision rule:** If an agent received new information, it is free to choose a new operation schedule. This new schedule might e.g. better fulfill the energy unit's constraints. If the new schedule leads to a better cluster schedule regarding the target schedule ζ, a new solution candidate has been found and is communicated.
- **Agent model:** COHDA is used for a planning task – once a set of possible operation schedules is determined, no monitoring or control of the energy unit is needed. Therefore, the agent model is deliberative.

3 DynaSCOPE

In the following, we will present the design concept that leads from COHDA to DynaSCOPE, a fully distributed algorithm for the **Dyna**mic **S**cheduling **C**onstraint **OP**timization for **E**nergy units. In table 2, the characteristics of COHDA are combined with the functional requirements for continuous energy planning as discussed in section 2.2. Extending COHDA to products leads to an extension need concering the most characteristics. The agent model though has not to be changed to fulfil this requirement.

Characteristic	Products	Incidents	Replanning	Costs	Constraints
Information exchanged	×			×	
Exchange criterion		×			
Exchange topology	×				
Information integration	×				
State evaluation	×		×	×	
Decision rule	×				×
Agent model		×			

Tab. 2: Extension needs by characteristics and functional requirements. ×: Extension or adaptation regarding this characteristic is needed to fulfil the respective requirement.

The guiding principle in setting up the design concept from these extension needs was to yield a fully functional system with known expected behavior in each step. By this means, a simulative evaluation can be performed for each extension, thus following the iterative

SGAE process model. To this end, the extension needs have to be restructured to a stepwise extension approach. In figure 1 the three-step design concept is shown. In design step 1, COHDA is extended to handling products and taking costs for schedules into account. Although this extension leads to a change in many characteristics (see table 2) a simple evaluation case can be defined for evaluation purposes: If only one product is put on the market, only one global DVPP will be formed. Thus, the algorithmic features can be directly compared to COHDA. The characteristics of the basic algorithm are documented in detail in [Hi14] – they serve as test base when evaluating the result of this design step. In design step 2, energy unit specific soft-constraints are introduced in the scheduling by extending the local evaluation of feasible operation schedules. A simulative approach with exemplary soft constraints can be used to show that these are factored in the planning process. The last design step leads to DynaSCOPE as fully distributed algorithm for continuous energy unit planning: Incidents are detected by the agents and a rescheduling is triggered if needed. The agents now have to reflect not only planning but control issues, as a continuous monitoring and control of the energy units is needed. The evaluation of the resulting software artefact reflects the requirements as discussed in section 1.

In the following, we will present some details of design steps 1 and 3, including problems by using a direct extension approach in step 1. Design step 2 is concerned with the integration of soft constraints and will be covered in a separate contribution.

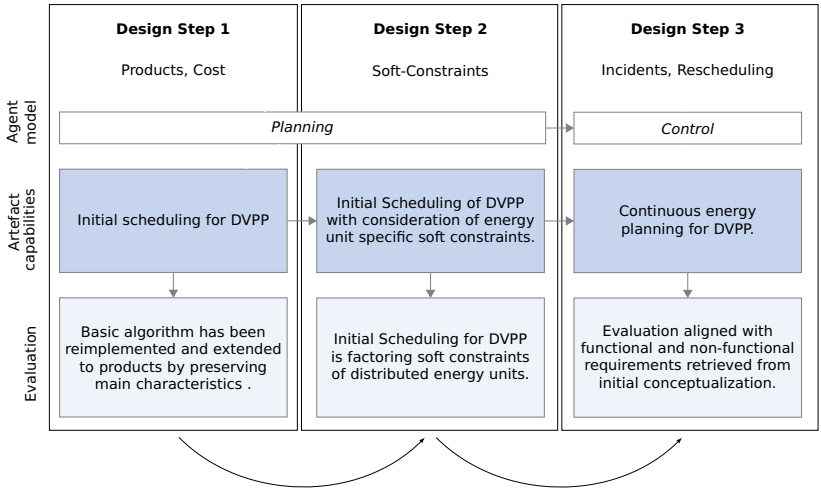


Fig. 1: Design Concept: Building DynaSCOPE from COHDA. [Ni15, translated from German]

4 Handling Products

Product specific coalitions of energy units are the main feature of DVPPs compared to static VPPs. It has to be decided how products are handled during agent information exchange, information integration, state evaluation and agent decision. Additionally it has

to be decided if the agents' exchange topology should change. In the following we will first present a straightforward approach to adapting COHDA to products and the problems resulting from this leading to some basic design decisions for DynaSCOPE.

For a first extension to products, the following extensions have been made:

- Information exchanged: The agents exchange their full operation schedule and add information on the product negotiated within the respective DVPP.
- Exchange topology: An exchange topology is setup within each DVPP.
- Information integration: The agents incorporate all knowledge they receive and add product specific information.
- State evaluation: The agents evaluate the operation schedules received depending on the DVPPs they belong to.

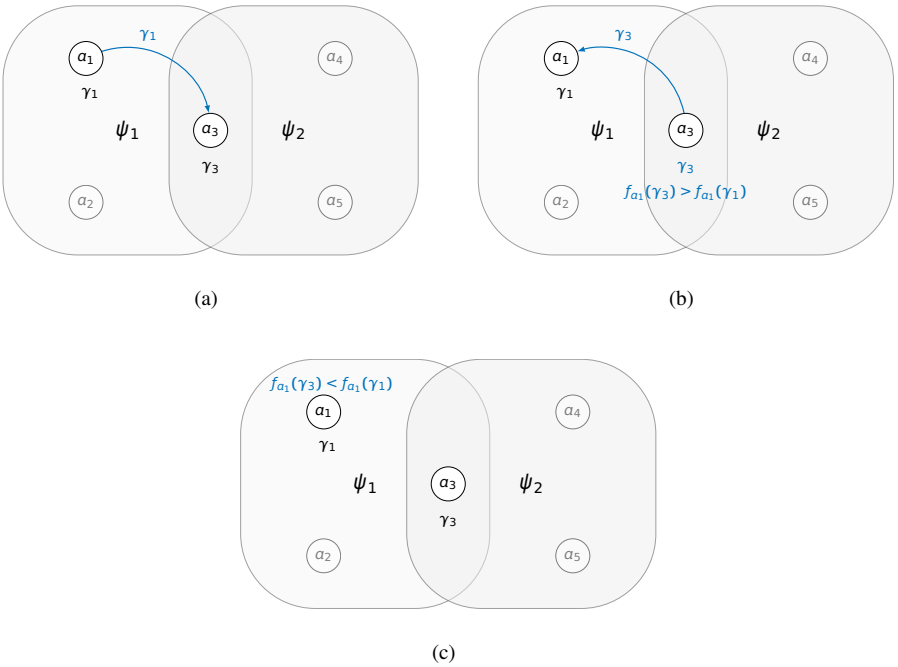


Fig. 2: Direct extension of COHDA to the product-specific DVPP concept.

In figure 2 a simple example is shown to illustrate why this straighforward extension approach is not feasible. In the example given, two DVPPs exist, with agent a_3 contributing to both DVPPs ψ_1 and ψ_2. Agents a_1 and agent a_3 both are aware of a solution candidate, i.e. a set of operation schedules for all agents within both DVPPs. In the first step (see figure 2(a)) agent a_1 sends its new solution candidate γ_1 to agent a_3. Agent a_3 now evaluates this solution candidate and compares its product performance to its own solution candidate γ_3. As agent a_3 contributes to both DVPPs, the solution candidate given by agent a_1 may perform worse in terms of product fulfillment than solution candidate a_3. Therefore, agent a_3 sends solution candidate γ_3 to agent a_1 (see figure 2(b)). Now agent a_1 evaluates

solution candidate γ_3 and compares its performance regarding DVPP ψ_1 to solution candidate γ_1. In the example given, γ_1 performs better regarding ψ_1. As no new information was incorporated or generated by agent a_1 the algorithm terminates. As both agents stick to different solution candidates though, the algorithm did not converge (see figure 2(c).

Using a simple example as depicted in figure 2 it could be shown that a straighforward approach for extending COHDA to DVPPs as product specific coalitions will sacrifice convergence. Some design decisions had to be taken as a result of this:

First, if products are the basis for DVPP, operation schedules should not cover the whole planning horizon (i.e. 24 hours). Therefore, only the product specific slot of an operation schedule, the so-called product schedule) should be communicated within a DVPP. The assigment of a product schedule to all agents within a DVPP is called DVPP schedule. Beside ensuring convergence, this can be seen as an aspect of data minimization. Second, the evaluation function for evaluation the solution candidate can no longer be global on a system level. Therefore, for evaluating new solution candidates, all agents take into account only the product performance within the DVPP under negotiation. In figure 3 the main ap-

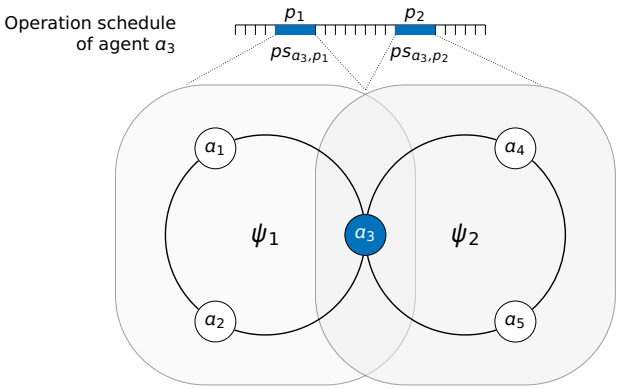


Fig. 3: Communication of product schedules as DVPP-specific excerpt of an operation schedule.

proach to reassure convergence of the algorithm is depicted using the example from figure 2: Although an energy unit has to fulfil a full operation schedule for the whole planning horizon (typically 24 hours), only the product specific excerpt is communicated within a DVPP. For the example given, agent a_3 communicated it's product schedule ps_{a_3,p_1} within DVPP ψ_1 and product schedule ps_{a_3,p_2} within DVPP ψ_2. This reveals some limitations regarding the usage of flexibility of the energy units but ensures convergence.

With this adaptations, the product specifics of DVPPs are integrated into COHDA.

5 Events and Rescheduling

As soon as agents have to detect if their respective energy units follow the negotiated operation schedule, the system has to evolve from an planning system to an agent based

control system. As discussed in [BS00] this has crucial influence the agent model needed to fulfil this task. Additionally, the process model has to be adapted to interleave the processes for planning, monitoring and control in an appropriate way. Both aspects are discussed in the rest of this section.

5.1 Agent Model

Software agents in real-world environments are usually categorized in three basic types that define their interaction with the physical world: Reactive agents directly transfer a sensoric input into an action. Reaction of this type of agents can be realized very fast. Deliberative agents model the information retrieved in an internal world model that is used for planning and proactive behavior. Hybrid agents include both aspects (see e.g. [We13] for a detailled discussion on different agent types).

For the task at hand, the continuous planning in DVPP, a hybrid agent model is needed: For the (day-ahead) planning process, a detailled world model, including the energy units capabilities is needed. For monitoring and control though, direct reaction to sensoric input should be possible to ensure a fast reaction to critical states. During simulative evaluation, the differences might not be crucial. When the system is transferred to the field though, the reactive parts might be realized on different (e.g. more reliable, faster and more costly) hardware.

With the InteRRaP architecture a hybrid architecture has been presented many years ago [MP94]. InteRRaP is a vertically layered architecture that represents both knowledge and behavior in different layers. From bottom to top, the layers pass from sensors and actors in the physical world to the cooperative agents' world. In figure 4, the DynaSCOPE agent model based on this InteRRaP architecture is depicted. It shows two interfaces – the unit interface at the bottom and the cooperation interface at the top. Behavioral layers and world model layers address different abstraction layers regarding physical and agent world. In the local evaluation layer, the sensoric input is evaluated to detect incidents: Does the energy unit follow the required operation schedule? Basic unit knowledge like the current operation schedule is needed in this layer. In the second layer, the local planning layer, the agent generates optional operation schedules for the energy unit. Different realizations are possible to model the local planning knowledge, i.e. the energy unit's flexibility. For DynaSCOPE, a vector-based surrogate model using a high-dimensional representation of possible unit states has been used as proposed by Bremer et al. in [BS13] called search space for the rest of this contribution.

For the cooperative planning process, two layers have been introduced in the DynaSCOPE agent model: The global planning layer holds behavior and information for all agents in the respective DVPP needed during the cooperative search of a new cluster schedule. The cooperation layer and knowledge are needed to define, which agents communicate during cooperative search of the new DVPP schedule.

A detailed description of all information stored at the different layers is given in [Ni15]; an excerpt is given in figure 4.

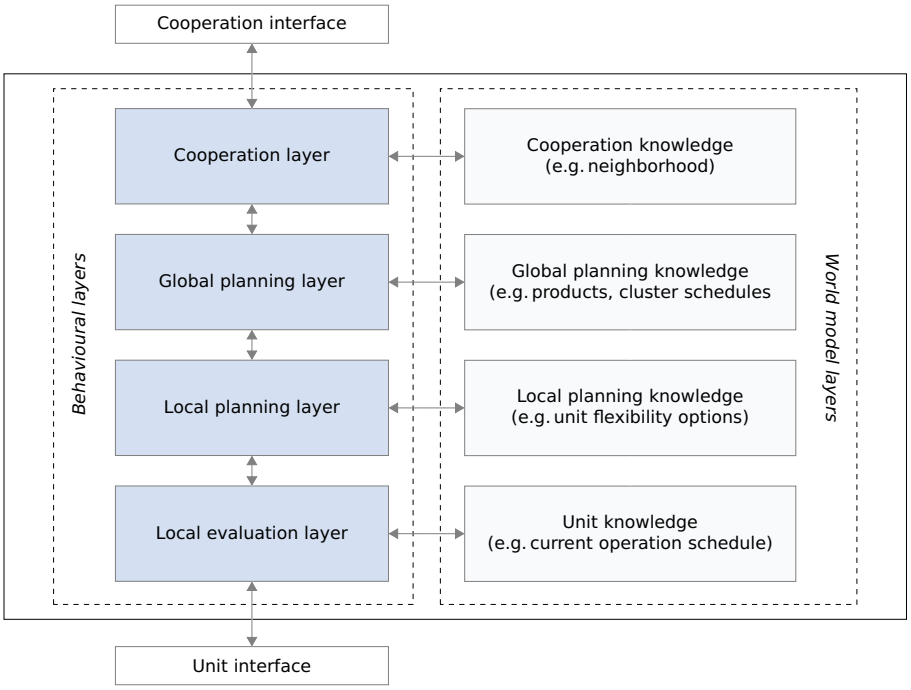


Fig. 4: Hybrid agent model for continuous energy planning

5.2 Process model

DynaSCOPE realizes the continuous energy planning within DVPPs in a fully distributed manner. The processes may be triggered by two main different events, thus comprising entry points to the DynaSCOPE agent behaviors:

- An agent detects an **incident** at its respective energy unit.
- An agent receives a **message** from an agent within the same DVPP.

In the following, we will describe the first entry point to DynaSCOPE. The second one, the process of cooperative search, is similar to the search process as defined for COHDA, although some adaptations have been made to include the incident information during information syntheses. A detailled description regarding this aspect is given in [Ni15].

In figure 5 the processes of detecting and processing an incident are shown. The figure reuses the agent model as shown in figure 4, but focusses on the behavior layers. The process starts at the local evaluation layer. The current operation schedule already has been transferred to the unit in a prior step not shown.

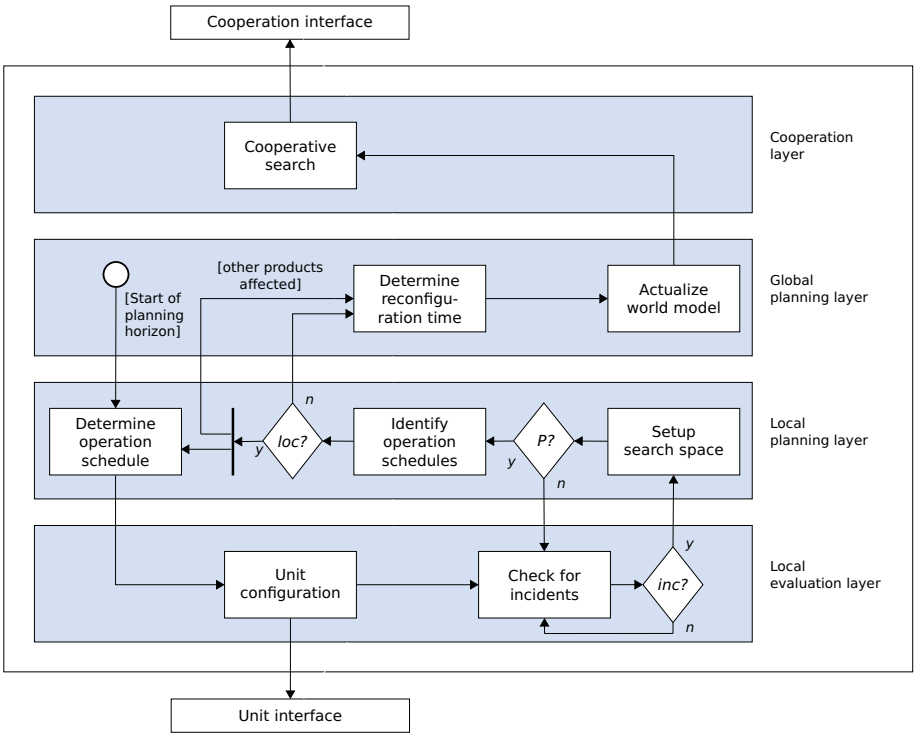


Fig. 5: DynaSCOPE processes from the perspective of an agent detecting an incident at its energy unit. *inc?*: Has an incident been detected? *P?*: Is a product affected by the incident? *loc?*: Is a local compensation possible?

1. Check for incidents: In each monitoring interval the agents check, if the operation schedule is fulfilled. If an incident has been detected, the local planning layer is activated.

2. Setup seach space: As the energy unit state changed, the flexibility model is no longer valid. A new search space is created based on the current energy units behavior.

3. Identify product affected: Not all incidents might affect a product. If no product is endangered by the incident, no further action is needed and the agent returns to the local evalaution layer.

4. Identify operation schedules: If a product is affected by the detected incident, the agent determines operation schedules using the new search space. Local soft constraints might be reflected during this step.

5. Check for local compensation: If the agent can identify a new operation schedule that compensates the incident locally, this operation schedule is passed to unit con-

figuration. If no local compensation is possible or additional DVPPs and thus products are affected, the global planning layer is called to action.

6. Determine reconfiguration time: As the agent has to start the cooperative search for a new cluster schedule, a reconfiguration time is set. The operation schedule defined by then will be used for unit configuration. The first possible reconfiguraton time is at the end of the current planning interval. With the typical time resolution of 15 minutes nearly 15 minutes can be used for the cooperative planning process.

7. Actualize world model: In the next step the agent adds the information on the detected incident and the invalid operation schedule to its world model.

8. Cooperative search: In the last step the agent sends the DVPP-specific parts of its world model to all agents within its neighborhood, thus restarting the cooperative search process. From then on, the agent will process messages received from other agents at the cooperation interface.

6 Evaluation

6.1 Experimental setup

The evaluation of DynaSCOPE has been done using a JAVA-based implementation of the algorithms. The vector-based search space model has been implemented by Jörg Bremer, University of Oldenburg. More details are presented in [Ni15]. For the results presented here, the following scenario has been chosen:

- Energy units: 20 CHPs (4.7 kW_{el}), 20 PV plants (10 kW_p), 10 heatpumps (5 kW).
- Product: 11 a.m. to 3 p.m., 150 kWh to be delivered each hour.
- Message delay: 200 ms.
- Unit flexibility model (search space) generated for a spring day.
- Incidents: CHP outages and PV prognosis deviations.

The allocation of incidents within the 4-hour product horizon is important for the evaluation of the overall system. Therefore, this aspect has been implemented in a deterministic but random based way: We choose different temporal allocations of incidents in a random way but use random seeds to guarantee reproducibility. 100 runs have been chosen for each scenario with a defined set of incidents allocated over time in such a way.

6.2 Incident detection and rescheduling quality

In figure 6 an example is given for a run with 10 CHP outages.[3] On the X axis, the solution evaluation is chronologically ordered: Each time an agents evaluates a DVPP schedule, this is logged within the system. On the Y axis, the expected product delivery performance

[3] The term outage may be misleading here: It can be understood as any type of unscheduled switching off.

is given. A value of 1.0 means perfect product delivery (i.e. 150 kWh in all 4 product hours). In the left part of the diagram, the initial planning process is shown: The agents yield an expected product delivery performace of 0.99 in the day ahead planning process. For all following values, the simulated time is given on top of the diagram for ease of understanding. Incidents are depicted using arrows.

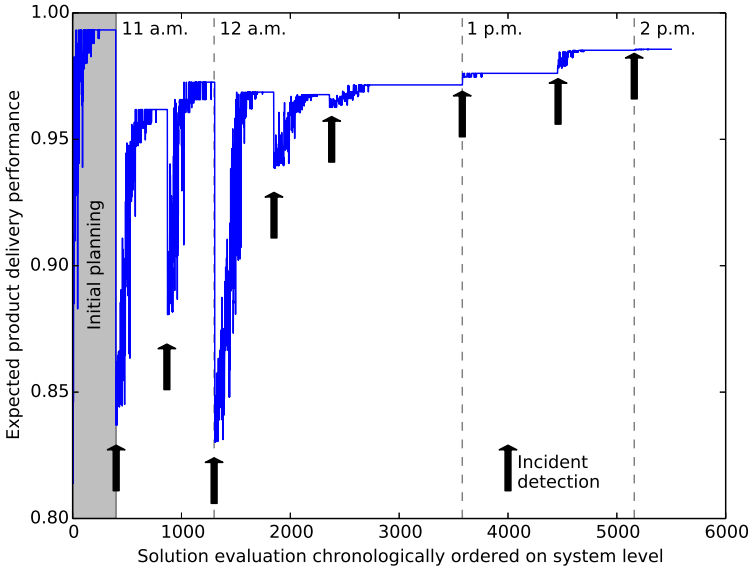


Fig. 6: Example for incident detection and expected product delivery performance.

In the first half of product delivery, each incident leads to a reduced product delivery performance first. The agents manage to enhance this value within the cooperative search for a new DVPP schedule though, thus enhancing the value. Convergence of the processes can be recognized from the plateaus right before the next incident. In the example given, the initial performance is not reached.

During the second half of product delivery, a very interesting effect is observed: with an incident, the product delivery performance does not decrease at all, but increases to a better value at once. This effect can be explained when examining the time span passed since begin of product delivery: The shorter the remaining product time span, the smaller is the expected effect of an outage.

In figure 7 the final product delivery performance is shown for all 100 runs per experimental setup and incidents configuration in a boxplot. The median is depicted with the blue line within the quartiles' box. The arithmetic mean is shown using a blue circle. The upper diagram shows the results for 0 to 10 CHP outages only, whereas the lower one shows the results of simulation runs with 4 PV prognosis deviations and 0 to 10 CHP outages combined. If no incidents are inserted, the final quality is about 0.99 for all runs (see upper left corner in figure 7). In both diagrams it can be seen, that the more incidents are inserted, the lower is the final performance. In all cases, the final product delivery performance exceeds

the expected values after incident detection shown in figure 6: DynaSCOPE leads to better product delivery in all simulation runs.

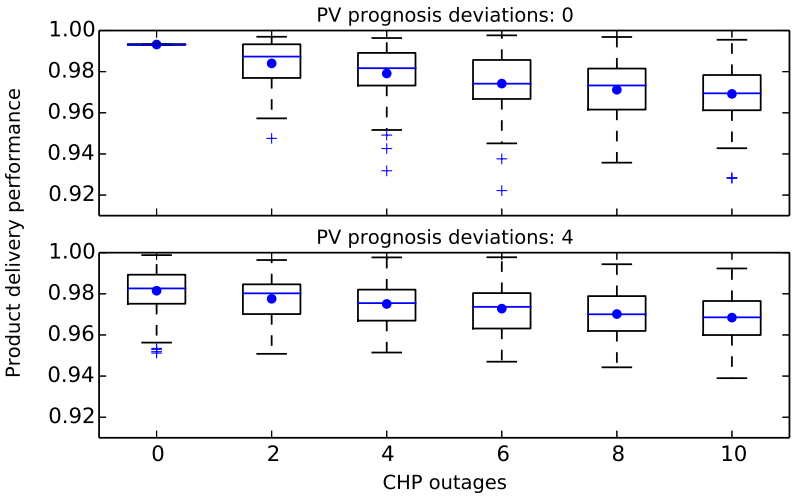


Fig. 7: Product delivery performance for a varying amount of incidents.

6.3 Communication Overhead

One aspect in evaluating the scalability of DynaSCOPE regarding the number of incidents detected in the field is the communication overhead in terms of mean number of messages sent by each agent. In figure 8 this value is set into relation with the number of incidents using boxplots. The same experimental setup was used as described in section 6.1. The maximum value can be found at about 550 with 14 incidents within the 4 hour product horizon. Furthermore, the number of messages grows in a sublinear manner. This can be explained from a DynaSCOPE detail not presented in detail in the work at hand: The more incidents are introduced in the 4-hour product horizon, the more incidents take place in the same planning interval. As the cooperative search processes are integrated by DynaSCOPE, the number of messages should raise sublinear with the number of incidents.

7 Conclusion and Future Work

In the work presented here, we introduced DynaSCOPE as a fully distributed continuous planning approach for decentralized energy units organized in DVPPs. DynaSCOPE has been developed by extending COHDA, a distributed optimization heuristic, to a distributed monitoring and control system. We evaluated DynaSCOPE using a simulative approach. It could be shown, that DynaSCOPE effectively enhances product delivery and thus allows for the delivery of an energy product with less reliable energy units prone to prognosis

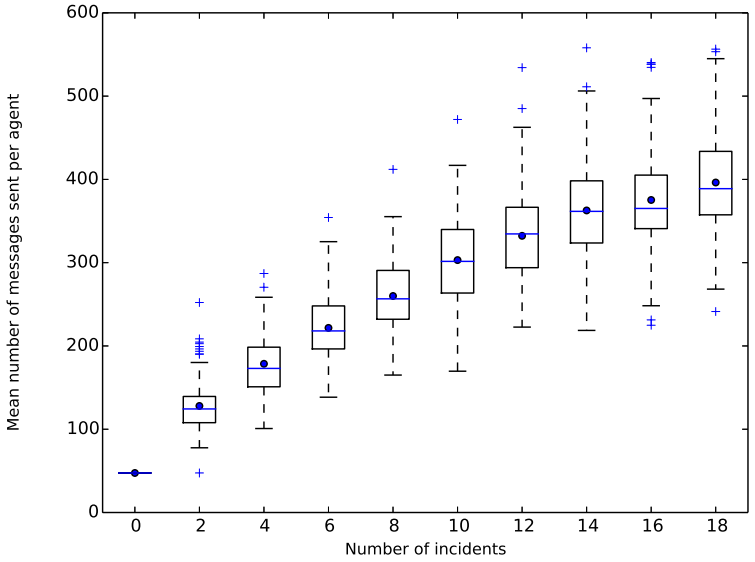


Fig. 8: Mean number of messages sent per agent by number of incidents inserted during product horizon.

deviations and outages. As a first distributed approach for this task following the DVPP concept, the performance of the system cannot be compared to other planning approaches yet. In the first iteration of Smart Grid Algorithm Engineering presented here, main findings have been retrieved, further motivating the extension of the system and narrowing the gap to an application in the field. In current work, we evolve DynaSCOPE within the open source energy unit aggregation, planning and control system Open VPP[4].

8 Acknowledgements

The Lower Saxony research network ‘Smart Nord’ acknowledges the support of the Lower Saxony Ministry of Science and Culture through the “Niedersächsisches Vorab” grant programme (grant ZN 2764).

References

[An10] Anders, Gerrit; Siefert, Florian; Steghöfer, Jan-Philipp; Seebach, Hella; Nafz, Florian; Reif, Wolfgang: Structuring and Controlling Distributed Power Sources by Autonomous Virtual Power Plants. In: Proceedings of the IEEE Power and Energy Student Summit (PESS). 2010.

[AYG96] Akkermans, Hans; Ygge, Frederik; Gustavsson, Rune: Homebots: Intelligent decentralized services for energy management. In: Fourth International Symposium on the Management of Industrial and Corporate Knowledge. Rotterdam, NL, 1996.

[4] http://openvpp.offis.de

[BS00] Bussmann, Stefan; Schild, Klaus: Self-Organizing Manufacturing Control : An Industrial Application of Agent Technology. In: Proc. of the 4th Int. Conf. on Multi-agent Systems (ICMAS'2000). Boston, MA, USA, pp. 87–94, 2000.

[BS13] Bremer, Jörg; Sonnenschein, Michael: Constraint-handling for optimization with support vector surrogate models. In: ICAART 2013. Barcelona, 2013.

[Ch11] Chapman, Archie C.; Rogers, Alex; Jennings, Nicholas R.; Leslie, David S.: A unifying framework for iterative approximate best-response algorithms for distributed constraint optimization problems. The Knowledge Engineering Review, 26(04):411–444, November 2011.

[Gu99] Gustavsson, Rune: Agents with Power. Communications of the ACM, 42(3):41–47, 1999.

[Hi14] Hinrichs, Christian: Selbstorganisierte Einsatzplanung dezentraler Akteure im Smart Grid. PhD thesis, 2014.

[HLS14] Hinrichs, Christian; Lehnhoff, Sebastian; Sonnenschein, Michael: A Decentralized Heuristic for Multiple-Choice Combinatorial Optimization Problems. In: Operations Research Proceedings 2012. Springer, pp. 297–302, 2014.

[Le10] Lehnhoff, Sebastian: Dezentrales vernetztes Energiemanagement - Ein Ansatz auf Basis eines verteilten Realzeit-Multiagentensystems. Vieweg + Teubner, 2010.

[MP94] Müller, Jörg P.; Pischel, Markus: An architecture for dynamically interacting agents. Journal of Intelligent and Cooperative Information Systems, 3(1):23–45, 1994.

[Ni12] Nieße, Astrid; Lehnhoff, Sebastian; Tröschel, Martin; Uslar, Mathias; Wissing, Carsten; Appelrath, Hans-Jürgen; Sonnenschein, Michael: Market-based self-organized provision of active power and ancillary services. In: Complexity in Engineering (COMPENG), 2012. Aachen, 2012.

[Ni15] Nieße, Astrid: Verteilte kontinuierliche Einsatzplanung in Dynamischen Virtuellen Kraftwerken. PhD thesis, 2015.

[NTS13] Nieße, Astrid; Tröschel, Martin; Sonnenschein, Michael: Designing Dependable and Sustainable Smart Grids - How to Apply Algorithm Engineering to Distributed Control in Power Systems. Environmental Modelling & Software, 2013.

[Po13] Pournaras, Evangelos: Multi-level Reconfigurable Self-organization in Overlay Services (PhD thesis). TU Delft, 2013.

[RRJ09] Rahwan, Talal; Ramchurn, Sarvapali D.; Jennings, Nicholas R.: An Anytime Algorithm for Optimal Coalition Structure Generation. Journal of Artificial Intelligence Research, 34(1):521–567, 2009.

[Ta09] Talbi, El-Ghazali: Metaheuristics: From Design to Implementation. John Wiley & Sons, Inc., 2009.

[Tr10] Tröschel, Martin: Aktive Einsatzplanung in holonischen Virtuellen Kraftwerken. OlWIR, Oldenburger Verl. für Wirtschaft, Informatik und Recht, Oldenburg, 2010.

[We13] Weiss, Gerhard: Multiagent Systems. MIT Press, 2013.

[ZV12] Zhabelova, Gulnara; Vyatkin, Valeriy: Multiagent Smart Grid Automation Architecture Based on IEC 61850/61499 Intelligent Logical Nodes. IEEE Transactions on Industrial Electronics, 59:2351–2362, 2012.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Userinteraktion und Energieoptimierung

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Userinteraktion und Energieoptimierung

Prof. Dr. Dominic Becking[1], Jan Budke[1]

Abstract: Dieser Workshop behandelt den Einfluss von Userinteraktion auf Technologien der Energieoptimierung - insbesondere, jedoch nicht ausschließlich Userinterfaces und Userinteraktion in intelligenten Gebäuden. Die hohe Bedeutung der Benutzerfreundlichkeit ergibt sich aus der Rolle des Nutzers als wichtigem Einflussfaktor auf Energieeinsparung im privaten Umfeld und darüber hinaus bei der Interaktion mit modernen Zweckbauten. Ein weiterer besonders interessanter Aspekt ist die Interdisziplinarität dieses Themas. Die Planung, Umsetzung und der Betrieb von Technologien zur Energieoptimierung betrifft nicht nur Informatiker, sondern ebenso Ingenieure, Architekten, Psychologen und Bauingenieure.

Die Beiträge des Workshops umfassen daher auch viele Themenbereiche. *Microcurricula in augmentierter Realität* beschäftigt sich mit der Akzeptanz von intelligenter Gebäudetechnologie. *Der moderne Brandschutz* legt den Fokus auf die Implementierung von Fluchtwegen, die im Brandfall aktiviert werden und die Menschen sicher zum nächsten Notausgang leiten sollen. Um gesundheitliche Aspekte geht es bei *Integration of a fall detection system into the intelligent building*. Das System soll Stürze erkennen und lokalisieren können und gleichzeitig möglichst diskret agieren. *Die Wirkung von Feedback und Goal Setting auf den Energieverbrauch* erstellt eine Analyse und Systematisierung von Studien und Pilotprojekten zum Thema Feedback beziehungsweise Goal Setting. Ziel ist es, den Energieverbrauch von Privathaushalten transparenter zu gestalten.

Keywords: Userinteraktion, userinterface, augmented reality, usability, human computer interaction, fall detection, Brandschutz, feedback, goal setting

Planungskomitee

Prof. Dr. Dominic Becking (FH Bielefeld, Informatik)
Prof. Dr. Matthias König (FH Bielefeld, Informatik)
Prof. Dr.-Ing. Oliver Wetter (FH Bielefeld, Elektrotechnik)
Prof. Dr.-Ing. Ulrich Schramm (FH Bielefeld, Architektur)
Prof. Dr. Sibylle Reichart (FH Bielefeld, Wirtschaftspsychologie)
Jan Budke (FH Bielefeld, Informatik)

[1] FH Bielefeld, Campus Minden, Ringstraße 94, 32427 Minden, [vorname.nachname]@fh-bielefeld.de

Beiträge

- Microcurricula in augmentierter Realität
- Der moderne Brandschutz
- Integration of a fall detection system into the intelligent building
- Die Wirkung von Feedback und Goal Setting auf den Energieverbrauch von Privathaushalten

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Microcurricula in augmentierter Realität – Ein Prototyp zur Hilfe bei der Interaktion mit intelligenten Gebäuden

Dominic Becking[1], Jan Budke[1] und Sybille Reichart[2]

Abstract: Neue und ungewohnte Bedienkonzepte stellen für die Benutzer oft ein Hindernis bei der Interaktion mit Technologie dar. Bei komplexen Technologien kommt es daher zu unerwünschten Nebenwirkungen. Durch Microcurricula, kurzen situations- und nutzerzentrierten Lerneinheiten, soll die Akzeptanz von intelligenten Gebäuden und das Wohlbefinden bei den Benutzern gesteigert werden. Für diesen Zweck wurde eine mobile App entwickelt, die Microcurricula in augmentierter Realität darstellt. Im Fokus stehen hierbei Userinterfaces zur Steuerung von intelligenter Gebäudetechnologie. Mit dem ersten Prototyp sollen vor allem die Effekte der App auf die Benutzer erfassbar gemacht werden. In einem Laborversuch wurden Testpersonen mit einem ihnen unbekannten Bediengerät konfrontiert und sollten mithilfe der App einige Funktionen des Gerätes nutzen.

Keywords: augmented reality, mobile learning, intelligent buildings, building automation

1 Einführung

Intelligente Gebäude sollen den Nutzern einerseits einen Komfortgewinn bieten und gleichzeitig den Energieverbrauch senken sowie einen Schutz vor kritischen Situationen bieten [Wil15]. Dafür werden verschiedene Überwachungs-, Steuerungs- und Regelungssysteme eingesetzt. Andrerseits werden durch diese Systeme wiederum neue Hürden für die Nutzer aufgebaut. Je komplexer die eingesetzte Technik ist, desto größer ist die Gefahr von unerwünschten Nebenwirkungen [Dic02].

Von besonderer Bedeutung sind daher die Usability und die Akzeptanz der Userinterfaces [Nie93]. Diese Bedeutung ergibt sich aus der Rolle des Nutzers als wichtigem Einflussfaktor auf Energieeinsparung im privaten Umfeld [Asc14] und darüber hinaus bei der Interaktion mit modernen Zweckbauten z.B. Bürogebäuden. Ihr volles Potenzial erreichen intelligente Gebäudetechnologien erst, wenn die Akzeptanz der Nutzer sichergestellt ist [Rus03]. Eine unzureichende Benutzerfreundlichkeit kann schnell zu einem höheren Energieverbrauch und damit auch zu erheblich höheren laufenden Kosten führen. Zudem werden die Technologien ohne den konsequenten Blick auf die Akzeptanz schnell zum Ärgernis für den Benutzer und können Vorbehalte, Ablehnung und Ängste hervorrufen [Bre72].

[1] Fachhochschule Bielefeld, Campus Minden, 32427 Minden, [vorname.nachname]@fh-bielefeld.de
[2] Fachhochschule Bielefeld, 33615 Bielefeld, [sybille.reichart]@fh-bielefeld.de

Widerstände gegen die Nutzung von energiesparenden Technologien können durch verbesserte HCI verhindert werden - jedoch müssen auch vorhandene Technologien einfach nutzbar gemacht werden.

Im Teilziel „Wohlergehen und Akzeptanz“ des Forschungsschwerpunkts InteG-F der Fachhochschule Bielefeld werden mögliche Barrieren für Wohlergehen und Akzeptanz intelligenter Gebäudetechnologien untersucht und Strategien zur positiven Beeinflussung von Gebäudenutzern erarbeitet. Der InteG-F wird gefördert durch das Ministerium für Innovation, Wissenschaft und Forschung des Landes Norrhein-Westfalen. Im Rahmen dieser Forschung wird eine mobile Applikation entwickelt, in der Microcurricula zu den Technologien mit Hilfe von augmentierter Realität präsentiert werden. Microcurricula sind kleine Lerneinheiten, die dem Nutzer die grundlegende Funktion und Bedienung eines Geräts erklären sollen.[3] Diese Lerneinheiten zu bestimmten Userinterfaces soll ein Nutzer über sein Smartphone abrufen können. Dafür erkennt die App automatisch um welches Userinterface es sich handelt und erhält die dafür vorgesehenen Informationen von einem Server. Art, Umfang und Sprache der Informationen hängen vom Profil des Nutzers ab. So soll sichergestellt werden, dass der Nutzer nur für ihn relevante, verständliche und interessante Informationen bekommt.

Der aktuelle Prototyp erkennt bereits Userinterfaces und zeigt dem Nutzer die dazu passenden Informationen an. Im Fokus der Entwicklung steht das Raumbediengerät UP 227 der Firma Siemens[4]. Grund dafür ist die Ausstattung des Neubaus auf dem Campus Minden der Fachhochschule Bielefeld mit diesen Geräten. Das Gebäude, das zur Zeit der Veröffentlichung kurz vor der Fertigstellung steht, beinhaltet die Bibliothek, die Mensa, und die Büro- und Seminarräume sowie Labore des Bereichs Technik.

2 Grundlagen

In der Forschung finden augmentierte Realität und mobiles Lernen eine große Beachtung. Olsson et al. [Ols13] untersuchten beispielsweise die Erwartungen und Bedürfnisse der Nutzer an mobile augmentierte Realität. Dabei ergab sich, dass die angezeigten Informationen vor allem gehaltvoll, lebhaft und kontextbezogen sein müssen. Dabei sind besonders die persönliche Relevanz und die Personalisierung von großer Bedeutung, da augmentierte Realität die persönliche Perspektive beeinflusst und die Sichtweise auf die Welt mitgestaltet.

Shirazi und Behzadan [Shi13] entwickelten eine Applikation für Smartphones, mit der Studenten technologischer Studiengänge ein zusätzliches Level an Informationen zu ihrem Lehrbuch erhalten können. Die App erkennt QR Codes und zeigt daraufhin Multimedia-Objekte wie z.B. Simulationen an, die zu einem tieferen Verständnis des

[3] Eine Begründung von Mikro- und Nanocurricula wird demnächst zur Diskussion gestellt.
[4] https://hit.sbt.siemens.com/RWD/app.aspx?ACTION=ShowProduct&KEY=5WG1227-2AB11&MODULE=Catalog, aufgerufen am 09.06.2015

Lernstoffes führen sollen. Die Mehrheit der Studenten befand diese Lernmethode als sehr effektiv und würde den Einsatz in weiteren Kursen begrüßen.

In einem anderen Versuch wurde eine digitale Karte mit textuellen Informationen und Expertenvideos mit einer App für augmentierte Realität verglichen. Über beide Ansätze sollten Studenten anhand eines bestimmten Bereichs des Campus mehr über Biodiversität und Nachhaltigkeit lernen. Ryokai und Agogino [Ryo] stellten fest, dass die Testgruppe, die augmentierte Realität benutzte, sich intensiver mit dem Thema beschäftigten. Die Testpersonen bewegten sich mehr und nahmen sich mehr Zeit für die Aufgabe. Augmentierte Realität schaffte demnach Anreize, selber aktiv zu werden und Dinge genauer zu betrachten. Dabei war es wichtig die Informationen kontextsensitiv und interaktiv anzubieten.

Die bisherigen Betrachtungsweisen zielen meist auf klassische Lernsituationen ab, wie man sie beispielsweise an Schulen oder in Universitäten vorfindet. Diese Arbeit untersucht hingegen, wie beiläufige Lernsituationen im alltäglichen Leben durch augmentierte Realität verbessert werden können. Es kann aber davon ausgegangen werden, dass augmentierte Realität bei Lernenden positiv aufgenommen wird [Shi13] [Ryo] [Dem]. Dabei ist es allerdings von Bedeutung, den Nutzer und seine Persönlichkeit zu berücksichtigen und auf die Relevanz der Information und den Kontext zu achten.

3 Der Prototyp

3.1 Entwicklungsziele

Der Prototyp soll zunächst grundlegende Funktionen erfüllen. Dazu gehört die Erkennung von Bildern (Markern), die mit einem spezifischen Userinterface verknüpft sind, so dass die dazugehörigen Informationen angezeigt werden können. Nachdem ein Bild erkannt wurde, werden die entsprechenden Microcurricula als dreidimensionaler Text eingeblendet. Abb. 1 zeigt eine schematische Darstellung der App. Das Raumbediengerät wurde erkannt und die dazu passende Lerneinheit wird angezeigt.

Für den ersten Prototyp wird auf eine Serverkomponente verzichtet. Sie ist für erste Tests nicht zwingend notwendig. Zudem wird das Nutzerprofil noch nicht automatisch berücksichtigt, sondern für einzelne Testdurchläufe jeweils statisch festgelegt. Der Prototyp soll zunächst Aussagen darüber ermöglichen, inwieweit Microcurricula in augmentierter Realität die Akzeptanz der Nutzer für Technologien steigern können und zu einer besseren Zusammenarbeit von Mensch und Technik führen. Dafür wird während der Konzeption und Entwicklung auch besonders darauf geachtet, dass der Prototyp selbst kein technologisches Hindernis für den Nutzer darstellt.

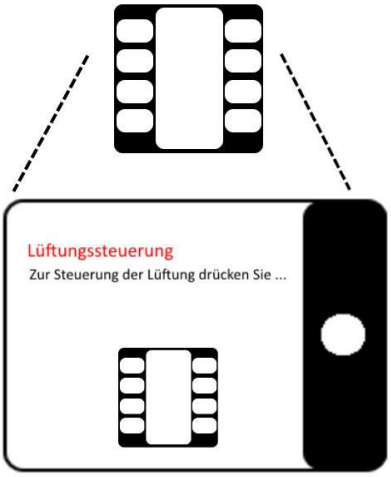


Abb. 1: Schematische Darstellung der App

3.2 Framework

Für die Entwicklung der App wird das Framework Vuforia[5] von Qualcomm eingesetzt. Mit dem Framework können Apps für mobile Geräte programmiert werden, die augmentierte Realität einsetzen. Es unterstützt die Plattformen Android, iOS und Unity 3D. Um Bilder und 3D Objekte in Echtzeit zu erkennen und zu verfolgen, wird eine computer-vision basierte Bilderkennung eingesetzt. Die Bilder müssen dabei nicht aus schwarzen und weißen Regionen bestehen, wie es bei QR- oder Bar-Codes der Fall ist wie z.B. noch bei [She02], sondern können beliebig gewählt werden. Das Framework erkennt automatisch die natürlich vorhandenen Merkmale der Marker. Damit 3D Objekte erkannt werden können, müssen diese zunächst mit dem Vuforia Object Scanner erfasst werden. Dieser erstellt eine Datei, in der die Objektmerkmale gespeichert sind [Qua151].

Zur Entwicklung des Prototypen wird Unity 3D[6] eingesetzt. Unity ist eine Laufzeit- und Entwicklungsumgebung des Unternehmens Unity Technologies, die die Entwicklung von Computerspielen und anderen interaktiven 3D-Grafik-Anwendungen erlaubt. Vuforia kann leicht als Plugin über den Unity Asset Store nachinstalliert werden. Zielplattformen sind neben mobilen Geräten auch PCs, Spielkonsolen und Webbrowser. Ein großer Vorteil von Unity ist, dass sie leicht zu benutzen ist. Bereits nach wenigen Minuten können erste Ergebnisse erzielt werden, ohne komplexen Code zu schreiben. Als Zielplattform wird Android gewählt. Aufgrund der Offenheit des Systems ist es

[5] https://www.qualcomm.com/products/vuforia
[6] http://unity3d.com/

einerseits weit verbreitet und andererseits existieren viele Frameworks und Schnittstellen, die das Zusammenspiel mit anderen Technologien ermöglichen.

3.3 Software

Die Serverkomponente wird vorerst durch eine einfache CSV Datei simuliert, in der die Texte der Microcurricula hinterlegt sind. Die App liest diese zu Beginn aus und kann so die zum Kontext passende Information jederzeit bereitstellen. Nach dem Start der App wird nach einem bekannten Marker gesucht. Wurde dieser erkannt, kann der Nutzer aus verschiedenen Lerneinheiten die für ihn relevante auswählen. Über Swipe-Gesten kann nun zwischen den einzelnen Schritten der Einheit gewechselt werden.



Abb. 2: Screenshoot der App

Für den ersten Prototypen werden nur die Hilfetexte für die grundlegenden Funktionen des Raumbediengeräts UP 227 hinterlegt. Es werden nacheinander die einzelnen Ebenen der Oberfläche des Bediengeräts beschrieben. Die oberste hat in diesem Fall keine Funktion und dient nur zur Visualisierung von Informationen. Über die unterste Ebene wählt man die gewünschte Funktionsgruppe, zum Beispiel Jalousie oder Licht. Die einzelnen Funktionen der Funktionsgruppe werden wiederum in der zweiten Ebene ausgewählt. Die dritte Ebene dient der eigentlichen Steuerung der Funktion. Zu den Texten der App werden außerdem noch Pfeile eingeblendet, die auf die jeweils beschriebene Ebene deuten (siehe Abb. 2).

4 Tests

Um zunächst die korrekte Funktionsweise der entwickelten App zu gewährleisten, wurden vor dem Laborversuch einige Funktionstests durchgeführt.

4.1 Funktionstests

Damit Programmierfehler erkannt und behoben werden können, wurde der Prototyp während der Entwicklung Funktionstests unterzogen. Dafür wurden die geschriebenen Methoden erweitert, so dass Logeinträge erzeugt werden. Eine Verbindung zu einem Server sowie die Auswahl einer Lerneinheit wurden zunächst nicht implementiert.

Während der Tests wurden verschiedene Kombinationen von Benutzereingaben getätigt. Dabei wurden auch solche berücksichtigt, die nicht zur aktuellen Situation passen und nicht zielführend sind. Anhand der Logeinträge wurde nachvollzogen, dass die Eingaben korrekt ausgewertet wurden und die erwünschten Funktionen ausgeführt wurden.

4.2 Laborversuch

Im Rahmen der Bachelorarbeit „Usability Testing – Augmented Reality Smartphone Applikation zur Erklärung der Gerätebedienung intelligenter Gebäudetechnik“ von Stefan Maldener wurde ein zweiteiliger Laborversuch mit jeweils zehn Testpersonen durchgeführt [Mal15]. Der Versuch bestand aus einem formativen Test, bei dem die Benutzer innerhalb eines realistischen Szenarios Aufgaben erfüllen sollten [Ric13]. Dafür wurde die Methode des lauten Denkens eingesetzt, bei der die Nutzer gebeten werden auszusprechen, was sie denken, erwarten und fühlen [Nie93].

Die Programmierung des Raumbediengerätes und die Erstellung und Verbesserung der App wurde dabei von Herrn Budke übernommen. Für die Konzeption und Durchführung des Versuchs sowie die Auswahl der Testpersonen war Herr Maldener verantwortlich. Die App wurde so konfiguriert, dass mit ihr die Handhabung des Raumbediengerätes UP 227 der Firma Siemens (siehe Abb. 3) erlernt werden kann. Als Ersatz für tatsächliche Aktionen in einem automatisierten Gebäude dienten als Rückmeldung auf Nutzereingaben beschriftete LEDs, die neben dem Bediengerät montiert wurden.



Abb. 3: Raumbediengerät UP 227

Den Testpersonen wurden fünf Aufgaben gestellt. Die Aufgaben lagen ihnen während des Versuchs schriftlich vor. Als einziges Hilfsmittel wurde ein Smartphone, auf dem die App installiert ist, bereitgestellt. Vor Beginn des Versuchs wurde die Bedienung der App kurz erklärt. Dabei ging es vor allem um die Ausrichtung des Smartphones zum Bediengerät und weniger um die Navigation innerhalb der App. Folgende fünf Aufgaben sollten bearbeitet werden:

1. Lesen Sie die Hilfetexte der App und machen Sie sich mit dieser vertraut.
2. Wählen Sie am Raumbediengerät eine Displayhelligkeit von 50 Prozent.
3. Schalten Sie am Raumbediengerät das Licht an der Tafel an. Als Bestätigung leuchtet auf dem LED-Board die dazu gehörige LED auf.
4. Senken Sie die Jalousie automatisch ab und fahren Sie diese danach wieder automatisch nach oben. Als visuelle Bestätigung dient Ihnen die gelbe LED „Jalousie runter/rauf“, die beim Absenken aufleuchtet und beim nach oben fahren erlischt.
5. Für eine PowerPoint-Präsentation wollen Sie verschiedene Funktionen nutzen, welche auf dem Gerät in einer programmierten Szene bereits miteinander verknüpft sind. Wählen sie die Szene „Präsentation“ an und aktivieren Sie diese. Bei Aktivierung leuchten 4 LEDs, nämlich „Jalousie runter“, „Beamer an“, „Leinwand runter“, und „Licht aus“ auf.

Abschließend wurden die Testpersonen interviewt. Dabei wurden ihnen folgende sechs Fragen gestellt und es gab die Möglichkeit eigene Anmerkungen einzubringen.

1. Hatten Sie vorher bereits Erfahrung mit ähnlichen Smartphone-Anwendungen?
2. Wie sinnvoll halten sie die Smartphone-App als Erklärung zur Gerätebedienung?
3. Würden Sie eine herkömmliche schriftliche Bedienungsanleitung bevorzugen?
4. Waren die Texte in der App für Sie verständlich formuliert?
5. Hatten Sie Probleme damit, die Funktion des Gerätes nach Betrachtung der App nachzuvollziehen?
6. Wie verständlich waren die Piktogramme des Raumbediengerätes für Sie?

Der zweite Testdurchlauf diente dazu, die Schlussfolgerungen aus dem ersten Test zu verifizieren und zu überprüfen, ob die daraus resultierenden Änderungen an der App zu dem gewünschten Ergebnis führen. Dafür wurden zunächst die Aufgabenbeschreibungen in die App integriert. Dadurch entfällt der Wechsel zu einem dritten Medium (Bediengerät, Smartphone, Aufgabenblatt). Insgesamt wurden die Hilfetexte kürzer gehalten und beziehen sich direkt auf die Aufgaben. Außerdem werden die Texte als Endlosschleife präsentiert, so dass ein Wechsel zwischen dem Ersten und dem Letzten möglich ist. Die Pfeile wurden zudem durch Kreise um die jeweiligen Tasten ersetzt.

4.3 Ergebnisse

Für die Quantifizierung der Ergebnisse wurden die benötigten Tastendrücke („Klicks“) bis zur Erfüllung der jeweiligen Aufgabe gemessen. Vor jeder Aufgabe wechselt das Bediengerät immer in den gleichen Ausgangszustand. Da für die erste Aufgabe keine Klicks notwendig sind, wird hier die benötigte Zeit gemessen.

Die Zeit für das Lesen aller Hilfetexte des ersten Tests liegt zwischen 60 und 170 Sekunden. Der Mittelwert liegt bei 120,5 Sekunden. Einige Probanden bemerken trotz Anzeige des Fortschritts nicht, dass sie den letzten Hilfetext erreicht haben.

Bei der zweiten Aufgabe liegt die durchschnittliche Klick-Zahl bei 9,2. Die Hälfte der Testpersonen benötigt lediglich die Mindestanzahl von vier Klicks. Beim zweiten Durchlauf steigt der Durchschnittswert auf 10,3 Klicks an. Der Anstieg kann zwei Ursachen haben. Zum einen fehlt dem Aufgabentext der Hinweis, in der App voranzuschreiten. Zum anderen drücken vier Testpersonen eine Taste der unteren Ebene des Bediengeräts zu lang, so dass sie in die Funktionsgruppe Licht wechseln. Beides führt zu Verwirrungen bei den Testpersonen und der Versuchsleiter muss teilweise eingreifen.

Für Aufgabe 3 soll ein Licht eingeschaltet werden. Die anschließende Befragung zeigte, dass die Testpersonen erwarteten, das Licht mit der rechten Taste einzuschalten, obwohl die linke Taste genutzt werden muss. Deshalb ist die durchschnittliche Klick-Zahl mit 19,3 hoch. Bis zum Benutzen der rechten Taste liegt das Mittel bei 7,9 Klicks. Da im zweiten Durchlauf die Texte angepasst wurden und Kreise die jeweilige Taste markieren, sinkt der Durchschnittswert hier auf 8,3 Klicks.

Aufgabe 4 benötigt mindestens sechs Klicks. Im Versuch liegt der Mittelwert bei 10,9 Klicks. Auch hier versuchen die Testpersonen die Jalousie über die rechte Taste zu schließen, wobei die linke Taste betätigt werden muss. Hinzu kommt, dass durch den Wechsel zwischen Bediengerät und App das Bediengerät zurück in den Ausgangsmodus wechselt und sich so die Klick-Zahl erhöht. Im zweiten Durchgang waren nur fünf Klicks notwendig, da die Jalousie nicht mehr hochgefahren werden musste. Hier konnte der Durchschnittswert auf 5,5 Klicks reduziert werden.

Bei der letzten Aufgabe zeichnet sich bereits ein Lerneffekt ab. Bei benötigten fünf Klicks liegt der Mittelwert bei 6,7 Klicks beziehungsweise bei 6,3 Klicks in der zweiten Testreihe.

Im anschließenden Interview stufen sechs Testpersonen der ersten Testreihe die App als sinnvoll ein. In der zweiten sind alle Testpersonen der Meinung, die App wäre sinnvoll. Sieben beziehungsweise neun würden sie sogar einem herkömmlichen Handbuch vorziehen. Die einzelnen Texte der App werden von 19 Testpersonen als verständlich bewertet. Nur ein Proband merkt an, dass durch die Unvertrautheit mit App und Bediengerät ein Nutzen der Texte sehr schwer fiele.

5 Fazit

Die Entwicklung des Prototyps und der anschließende Test im Labor sollten zum einen zeigen, ob Microcurricula in augmentierter Realität einen positiven Einfluss auf die Akzeptanz der Nutzer von intelligenter Gebäudetechnologie haben. Zum anderen sollten die grundlegenden Funktionen der App, das Erkennen von Markern und das Anzeigen der passenden Hilfetexte erprobt werden. Durch die Befragung der Testpersonen zeigt sich deutlich, dass die App sehr gut angenommen wird. Fast alle Testpersonen sprachen sich für die Benutzung aus und bevorzugen die App gegenüber einer üblichen Bedienungsanleitung. Die Tests zeigen auch, dass die Erkennung von Markern schnell und zuverlässig erfolgt, so dass ein alltäglicher Einsatz möglich ist.

Der Vergleich der zwei Testreihen im Labor macht auch deutlich, dass kurze, sehr präzise Informationen zu einem besseren Verständnis führen. Ein gutes Beispiel hierfür sind der Hinweis auf das Einschalten des Lichts auf der linken Seite und die Markierung der entsprechenden Tasten durch Kreise anstatt durch einen Pfeil. Diese Erkenntnis ist auch sehr interessant für die automatische Zuordnung von Profilen und Nutzern, die im weiteren Verlauf der Forschungen entwickelt werden soll.

Abschließend lässt sich feststellen, dass der Prototyp die Erwartungen erfüllt. Für die weitere Entwicklung ist es wichtig, weiterhin darauf zu achten die App so zu gestalten, dass sie keine zusätzliche Barriere für den Nutzer darstellt.

Literaturverzeichnis

[Asc14] Aschendorf, B.: Energiemanagement durch Gebäudeautomation, Springer, 2014.

[Bre72] Brehm, J.: Responses to loss of freedom: A theory of psychological reaktance, General Learning Press, Morristown, 1972.

[Dic02] Dickenberger, D.; Gniech, G.; Grabitz, H.J.: Die Theorie der psychologischen Reaktanz. In (Frey, D.; Irle, M., Hrsg.): Theorien der Sozialpsychologie, Verlag Hans Huber, Seattle, 2002.

[Mal15] Maldener, S.: Usability Testing – Augmented Reality Smartphone Applikation zur Erklärung der Gerätebedienung intelligenter Gebäudetechnik, Studiengang Wirtschaftspsychologie (B.Sc.), Fachhochschule Bielefeld Fachbereich Wirtschaft und Gesundheit, 2015.

[Nie93] Nielsen, J.: Usability Engineering, Academic Press, San Diego, 1993.

[Nol14] Noll, C.; Häussermann, B.; von Jan, U.; Raap, U.; Albrecht, U-V.: Demo: mobile augmented reality in medical education: an application for dermatology. In (ACM, Hrsg.): MARS '14., S. 17-18, 2014.

[Ols13] Olsson, T.; Lagerstam, E.; Kärkkäinen, T.; Väänänen-Vainio-Mattila, K.: Expected user experience of mobile augmented reality services: a user study in the context of shopping centres. In (Arnich, B.; Osmani, V.; Bardram, J.E., Hrsg.): Personal and Ubiquitous Computing., S. 287-304, 2013.

[Qua15] Augmented Reality (Vuforia), https://developer.qualcomm.com/mobile-development/add-advanced-features/augmented-reality-vuforia, Stand: 12.05.2015.

[Qua151] Vuforia Object Scanner, https://developer.vuforia.com/library/articles/Training/Vuforia-Object-Scanner-Users-Guide, Stand: 03.06.2015.

[Ric13] Richter, M.; Flückiger, M.: Usability Engineering kompakt, Springer, Berlin, 2013.

[Rus03] Ruser, H.; Borodulkin, L.; Leisner, D.: Multi-modal ‘Smart Home’ user interface. In: Signals, Systems, Decision and Information Technology, S. 26-28, 2003.

[Ryo] Ryokai, K.; Agogino, A.: Off the Paved Paths: Exploring Nature With a Mobile Augmented Reality Learning Tool. In (Lumsden, J., Hrsg.): The International Journal of Mobile HCI., S. 21-49, 2013.

[She02] Shelton, B.E.; Hedley, N.R.: Using Augmented Reality for Teaching Earth-Sun Relationships to Undergraduate Geography Students. In (IEEE, Hrsg.): Proceedings of the First IEEE International Workshop on ARToolKit. 2002.

[Shi13] Shirazi, A.; Behzadan, A.H.: Technology-enhanced learning in construction education using mobile context-aware augmented reality visual simulation. In (Pasupathy, R.; Kim, S.H.; Tolk, A.; Hill, R.; Kuhl, M.E., Hrsg.): Winter Simulation Conference., S. 3074 - 3085, 2013.

[Wil15] Wilson, C.; Hargreaves, T.; Hauxwell-Baldwin, R.: Smart homes and their users: a systematic analysis and key challenges. In (Thomas, P., Hrsg.): Personal and Ubiquitous Computing., S. 463–476, 2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Der moderne Brandschutz – Fluchtwege der Zukunft

Andreas Sawatzki[1] Uwe Sedlak[1] Oliver Wetter[1] Uwe Weitkemper[1] Martin Hoffmann[1]

Abstract:

Unter dem Begriff moderner Brandschutz versteht der Forschungsschwerpunkt „Gebäudetechnologien unter einem Dach“ (kurz: InteG-F) der Fachhochschule Bielefeld eine Koppelung des Brandschutzes mit der Gebäudeautomation, um die Vorteile beider Bereiche zu vereinen und das Potential des Brandschutzes zu maximieren. Daher wurde in einem interdisziplinären Team ein Modellhaus entwickelt, das genau diesem Zweck dienen soll und Forschungen an der Kombination der beiden Themen zu ermöglichen. Ein erster Ansatz ist die Implementierung von Fluchtwegen, die im Brandfall aktiviert werden und die Menschen sicher zum nächsten Notausgang leiten. Dazu werden Information über Ort des Brandes und Anzahl der Personen berücksichtigt und von der Gebäudeautomation bereitgestellt. In diesem Bericht wird das Modellhaus und seine Funktionen dokumentiert und ein erster Einblick in den modernen Brandschutz gezeigt. Zudem werden nach einer Betrachtung des Stands der Technik zukünftige Forschungsfragen aufgeworfen.

Keywords: Brandschutz, Gebäudeautomation, InteG-F, Fluchtwege

1 Einleitung

Das Modellhaus des Campus Minden der Fachhochschule Bielefeld, demonstriert zukünftige Entwicklungen im Bereich des Brandschutzes in intelligenten Gebäuden. Anlagen des technischen Brandschutzes wie beispielsweise Rauchmelder und Leuchtschilder sind bereits seit Jahrzehnten in zahlreichen Gebäuden vorgeschrieben. Um zukünftigen Anforderungen an Energieeffizienz und Komfort entsprechen zu können, werden immer mehr Neubauten mit Technik zur Gebäudeautomation ausgestattet. Diese Anlagen regeln die z.B. Beleuchtung und das Klima im Gebäude in Abhängigkeit von der Sonneneinstrahlung und Außentemperaturen. Es wird erwartet, dass diese technischen Systeme des Brandschutzes und der Gebäudeautomation zukünftig zusammenwachsen und gemeinsame Schnittstellen benötigen. Bisher bestehen keine standardisierten Schnittstellen. Dies hat zur Folge, dass für jedes neu errichtete Gebäude die Schnittstelle zwischen Gebäudeautomation und technischem Brandschutz von der Feuerwehr abgenommen werden. Dieser Prozess ist kostenintensiv und für Bauvorhaben ein Risiko. Das Modellhaus der Fachhochschule Bielefeld veranschaulicht eine mögliche Umsetzung dieser Schnittstellen, die als Basis für zukünftige Standards dienen soll. Anhand realistischer Szenarien, die aus einem Brandschutzkonzept abgeleitet wurden, steuert die Gebäudeautomation (über KNX-Bus) eine Fluchtweganzeige. Im Gebäude sind Rauch- und Präsenzmelder verbaut. Im Brandfall

[1] Fachhochschule Bielefeld, Fachbereich Campus Minden, Ringstraße 94, 32427 Minden, andreas.sawatzki@fh-bielefeld.de, uwe.sedlak@fh-bielefeld.de, oliver.wetter@fh-bielefeld.de, uwe.weitkemper@fh-bielefeld.de, martin.hoffmann@fh-bielefeld.de

kann diese Information verwendet werden, um das Gebäude besonders effizient zu evakuieren. Das System wurde im Rahmen eines interdisziplinären Forschungsprojekts unter Beteiligung von Bauingenieuren, Architekten, Informatikern und Elektrotechnikern entwickelt. Es basiert auf realistischen Annahmen bezüglich der rechtlichen Anforderungen an den Brandschutz und verbindet diese mit aktueller Technologie der Gebäudeautomation.

2 Aufbau des Modellhauses

Bei dem Demonstrator handelt es sich um ein Modell eines intelligenten Multifunktionsgebäudes im Maßstab 1:15. Einzelnen Etagen werden verschiedene Funktionen zugewiesen. So gibt es im Erdgeschoss einen Empfangsbereich mit Sitzmöglichkeiten sowie eine Cafeteria. Das erste Obergeschoss bietet eine offene Bibliothek, ein Archiv und den unteren Teil eines zweigeschossigen Audimax. Der obere Teil des Audimax befindet sich im zweiten Obergeschoss sowie diverse Arbeitstische. Auf dem letzten Obergeschoss ist eine Büroebene abgebildet. Dort sind zahlreiche Einzel- und Gruppenbüros platziert.

Wichtige Aktoren des Systems des Modellhauses sind Taster, Schaltaktoren und Rauchwarnmelder. Mittels Taster können bestimmte Szenen aktiviert werden (wie etwa Normalbetrieb oder die Simulation eines Brandes). Letzteres ist wichtig, da vermieden werden soll, dass das Modellhaus einem echten Brand ausgesetzt wird. Der Schaltaktor besitzt potentialfreie Kontakte, die als Identifikation der Szene genutzt werden. Die Konfiguration des Schaltaktors wird binär interpretiert und aktuell werden vier dieser Kontakte verwendet um bis zu 2^4 verschiedene Szenen zu verwalten.

Das Modell selbst ist modular aufgebaut. Grundsätzlich besteht das Haus aus zwei Hälften. Zudem ist es möglich, jede Ebene zu Demonstrationszwecken einzeln herauszunehmen und bei Bedarf auszutauschen. Außerdem besitzt jede Etage eine eigene Steuereinheit für die Kernlogik der Fluchtwegleitung (im Folgenden „Smart Escape Route Assistance“ oder SERA s. Abs 5). Zwischen der Decke einer Ebene und dem Boden einer anderen ist jeweils Raum für Verkabelungen sowie die Allgemeinbeleuchtung der darunterliegenden Ebene. In den Boden der Etagen sind RGB-LEDs eingearbeitet (siehe Abbildung 1). Diese visualisieren die Fluchtwege und leiten im Brandfall zum nächsten Notausgang. Eine weitere modulare Eigenschaft des Modells ist die Tatsache, dass die Etagen mit nur zwei Steckern verbunden werden, dazu gehören die Spannungsversorgung sowie die Datenleitung. Über diese werden die Informationen vom KNX-Bus an die Fluchtwegleitung weitergeleitet.

3 Funktionen

Das Modellhaus befindet sich in einem frühen Ausbaustadium. Das heißt, aktuell wurde neben dem Konzept und Bau des Hauses nur eine wichtige Funktion implementiert: Die integrierten LED-Streifen auf allen Etagen des Modells dienen einer sicheren Navigation zum nächsten Notausgang im Brandfall und sind eine verbesserte optische Alarmierung für beispielsweise gehörlose oder nicht ortskundige Personen. Um diese Funktion zu gewährleisten, wurde SERA entwickelt, welches die Verwaltung der LEDs übernimmt

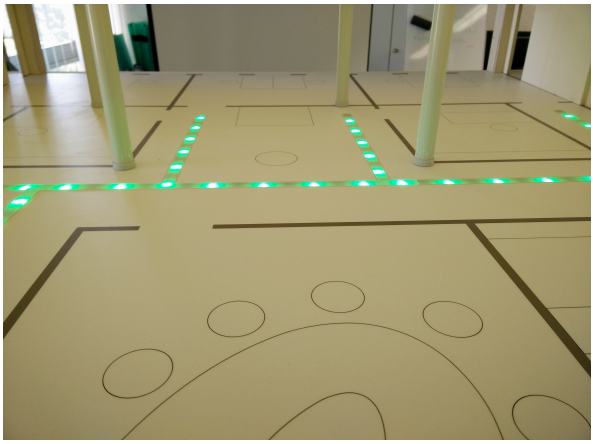


Abb. 1: LED-Streifen für Visualisierung der Fluchtwege

(s. Abbildung 2). Diese erhalten ein statisches Signal von dem KNX-Schaltaktor, interpretieren es und veranlassen die LEDs anschließend das entsprechende Muster darzustellen. Dabei sind die Platinen in Reihe geschaltet und die Signale der Sensoren werden von Ebene zu Ebene weitergereicht. Bei dem Signal handelt es sich um potentialfreie Kontakte, die entweder offen oder geschlossen sind. Somit können 2^4 verschiedene Kombinationen erreicht werden, da SERA aktuell vier Anschlüsse verwaltet. Durch diese Konstruktion ist die Fluchtwegleitung völlig von dem Rest des Systems losgelöst und autark. Es muss lediglich die Schnittstelle definiert werden, um die anliegenden Signale korrekt interpretieren zu können. Somit ist sichergestellt, dass SERA unabhängig von der Gebäudeautomation funktioniert, solange die eingehenden Signale definiert sind. Zukünftig ist eine andere Vernetzung von SERA denkbar.

Eine weitere Funktion, die aktuell bearbeitet wird, ist die Anbindung an bestehende Brandmeldeanlagen und die Implementierung des Kommunikationsstandards (nach VdS 2496). Dabei werden ebenfalls potentialfreie Kontakte verwendet, um sowohl Lösch- und Störungssignale der BMA zu kommunizieren als auch wechselwirkend Informationen von SERA zur Brandmeldeanlage zu übermitteln. Dadurch ist es möglich, meldegruppen- oder gar meldergenau zu bestimmen, wo ein Brand erkannt wurde und diese Informationen für die Visualisierung der Fluchtwege zu verwenden.

4 Stand der Technik

Im Falle eines Brandes ist es in erster Linie wichtig, die Personen sicher zum nächsten Notausgang zu leiten. Aber auch eine Aufteilung von großen Menschenmengen könnte im Extremfall ebenfalls lebenswichtig sein. Dazu muss das System allerdings wissen, wo sich die Personen befinden und wie viele Menschen sich dort aufhalten, um die Menge gegebenenfalls zu trennen und schneller zu evakuieren. Im Folgenden werden bekannte Methoden

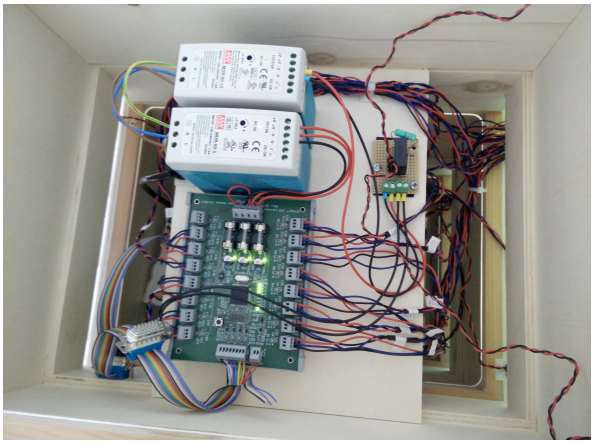


Abb. 2: Smart Escape Route Assistance – SERA

und Techniken aufgeführt, die sich zum Einen mit Brandschutz und zum Anderen mit der Personenzählung befassen.

4.1 Fluchtweglenkung

Der Brandschutz sieht vor, Fluchtwege in öffentlichen Gebäuden zu kennzeichnen. Dazu müssen Wegweiser an Decken und Wänden montiert werden sowie Fluchtwegpläne ausgehängt werden. Hinzu kommt, dass Notausgänge kenntlich gemacht werden und jeder Zeit passierbar sein müssen. Abbildung 3 zeigt ein Fluchtwegschild, wie es zum Beispiel in Hotels verwendet wird. Diese und weitere Schilder sollen die Menschen im Brandfall aus dem Gebäude leiten. Aufgrund der Tatsache, dass die Piktogramme gut sichtbar an Decken und Wänden montiert werden müssen, kann es vorkommen, dass sie bei starker Rauchbildung innerhalb großer Nutzungseinheiten oder in Bereichen öffentlicher Verkehrsanlagen nicht mehr gesehen werden können.

Deshalb beschäftigen sich Unternehmen wie beispielsweise Hanning & Kahl [HA15b] mit intelligenten Leitsystemen. Hanning & Kahl entwickelt „GuideLight“ [HA15a], welche eine intelligente Lösung für Leitsysteme und Informationssysteme im öffentlichen Personen Nahverkehr (ÖPNV) darstellt. „GuideLight“ kann analog zu SERA Menschen im Notfall zum nächsten Notausgang leiten. Zudem soll es Informationen vermitteln, wie etwa genaue Haltebereich von Zügen, bestimmte Abschnitte (z.B. Mehrzweckabteil) kenntlich machen oder durch blinkende Lichtelemente vor vorbeifahrenden Zügen warnen.

4.2 Personenzählung

Eine nützliche Information für den Brandschutz ist das Wissen über Anzahl und Position aller Personen im Gebäude. Kameras sind eine Möglichkeit, um Personen zu detektie-



Abb. 3: Fluchtwegschild
aus: [Bo15]

ren. Es gibt verschiedene Techniken, um Menschenmengen zu erfassen und messen. Dazu gehören Standardkameras, Infrarotkameras und Wärmesensorkameras.

Am Markt existieren bereits einige Lösungen zur Personenzählung in Kaufhäusern und im ÖPNV, jedoch nicht für den Brandschutz. Diese werden aktuell von spezialisierten Firmen verwenden, um dieser Aufgabe nachzukommen. Standardkameras werden zum Beispiel von den Firmen Visapix [VI15] und Cognimatics [Co15] vertrieben und zählen Menschen in Kaufhäusern bzw. im ÖPNV. Beide Firmen benutzen normale Kameras in kleinen Netzwerken und sammeln die Informationen der ausgewerteten Bilder auf Serversystemen. Die zweite Kameraart sind Wärmekameras, wie sie von der Firma LASE [LA15] verkauft werden. Bei diesen beiden Kameraarten werden Methoden und Techniken der Bildverarbeitung und Mustererkennung eingesetzt, um die aufgezeichneten Bilder zu analysieren.

Eine weitere Möglichkeit, um Menschen zu zählen, sind Infrarotkameras. In der Industrie werden diese unter anderen von den Firmen DILAX [DI15], Iris [In15] und MG [MG15] vertrieben. Diese Kameraart unterscheidet sich von den Standardkameras, da nicht die Personen in den Bildern gezählt werden, sondern die Bewegungsrichtung in beispielsweise Türen erfasst wird. Passiert ein Mensch eine Tür, in der eine Infrarotkamera verbaut ist, nimmt die Kamera eine Bewegung war und ermittelt anschließend die Richtung. Somit weiß das System, ob die Person den Raum betritt oder ihn verlässt.

Neben den Kameras können die Personen dem System auch selbst mitteilen, wo sie sich gerade befinden. Dazu besteht entweder die Möglichkeit, drahtlose Funktechnik wie RFID, NFC oder Bluetooth zu verwenden. Bei der drahtlosen Funktechnik ist es erforderlich, dass an diversen Stellen eines Gebäudes entsprechende Lesegeräte installiert werden (z.B. an Türen oder in Räumen). Die Reichweiten der Lesegeräte sind abhängig von der verwendeten Technik. So können NFC-Lesegeräte Sender bis zu zehn Zentimetern erfasst werden [WE15], Bluetooth und RFID hingegen kommen auf bis zu 100 Meter Reichweite [RF15, Bl15].

Alternativ besteht die Möglichkeit die Distanz zu WLAN Access Points (WAP) zu berechnen und diese Informationen an die Gebäudeautomation zu senden. Diese Methode setzt voraus, dass in dem Gebäude diverse WAPs montiert sind. In dem Fall können die WAPs abgefragt werden, wie viele Personen sich speziell an einzelnen Access Points angemeldet haben und somit eine konkrete Anzahl von Personen in einem Bereich bestimmt werden.

5 Ziele

Die primären Ziele, die mit Hilfe des Modellhauses erreicht werden sollen, sind solche, die den Menschen im Brandfall unterstützen. Vor allem assistierende Funktionen werden entwickelt und am Modell veranschaulicht. So soll zum Beispiel eine intelligentere Lösung für die Generierung der Fluchtwege implementiert werden. Diese sollen in Zukunft dynamisch berechnet werden, indem Informationen wie Brand- und Personenpositionen berücksichtigt werden. Sollte ein Brand direkt zwischen einem Notausgang und der Position einer oder mehrerer Personen entstehen, würde das aktuelle System den kürzesten, jedoch nicht zwingend den sichersten Weg dorthin darstellen. Die Fluchtwege werden momentan genau so visualisiert, wie die statischen Fluchtwegpläne aus Hotels und anderen öffentlichen Gebäuden. Dieses Problem soll umgangen werden, indem die Position des Feuers erkannt und ein alternativer Weg ermittelt wird, auch wenn dieser womöglich länger ist.

Ebenfalls denkbar ist eine direkte Kommunikation mit Aufzügen. Fahrstühle sind in der Regel so konfiguriert, dass sie im Brandfall ins Erdgeschoss fahren und dort mit geöffneten Türen verbleiben, bis der Brand gelöscht ist und sie wieder freigegeben werden. Falls ein Feuer allerdings im Erdgeschoss entsteht, wäre es fatal, wenn der Aufzug dort zum Stehen kommt, wenn sich noch Fahrgäste im Lift befinden. Eine mögliche Alternative könnte zum Beispiel sein, dass dem Aufzug mitgeteilt wird, wo sich das Feuer befindet und demnach eine andere Etage angesteuert wird.

Ein weiteres Ziel liegt im Bereich der Personendetektion, um große Personenmassen zu trennen und mögliche Personenstaus zu verhindern. Dies kann vor allem in Bestandsbauten wichtig sein, weil die vorhandenen Fluchtwegbreiten zu klein sein können und der Fluchtweg damit ggf. nicht ausreichend ist. Für getrennte Personenströme reicht ein Fluchtweg möglicherweise eher aus. Deshalb sollte bei der Entwicklung von SERA die mögliche Personenanzahl pro Fluchtwegabschnitt als Parameter definiert werden. Als weiterer Parameter sollte die „Qualität“ des Fluchtweg beurteilt werden. Dazu gehören Kriterien wie Brandlasten im Fluchtweg, Übersichtlichkeit des Fluchtwegs oder Stolperstellen im Fluchtweg. Diese könnten durch eine Checkliste oder ein kleines Tool aufgefasst und bewertet werden.

Der erste Ansatz zur Personendetektion bezieht sich auf die Verwendung von Wireless Access Points. Abbildung 4 zeigt die Grundrisse des ersten und zweiten Obergeschosses sowie eine mögliche Verteilung der WAPs. Die Wahl der Positionen für die WAPs wird in künftigen Untersuchungen evaluiert und möglicherweise optimiert, sodass eine bessere und genauere Bestimmung der Position möglich ist.

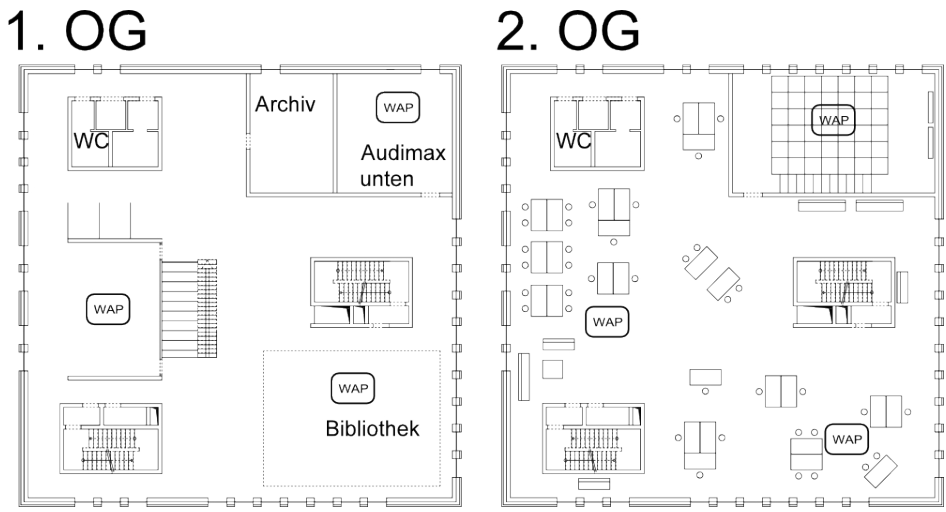
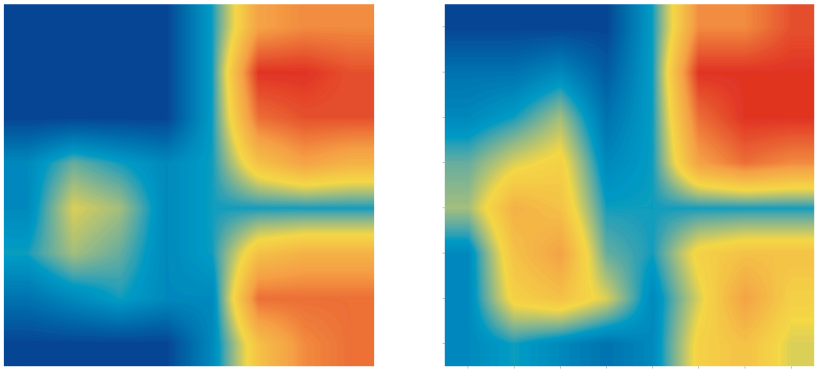


Abb. 4: Grundrisse und mögliche WAP-Verteilung

Um dieses Feature zu realisieren, muss eine Softwarekomponente entwickelt werden, welche die Informationen der WAPs auswertet und ein Gesamtbild über die Verteilung der anwesenden Personen erstellt. Abb. 5a und 5b zeigen eine mögliche Darstellung in Form einer Heatmap. Rote Bereiche repräsentieren viele Menschen und blaue hingegen bedeuten, dass sich keine Personen an diesem WAP angemeldet haben. In Abbildung 5a halten sich sehr viele Menschen im Audimax und in der Bibliothek auf, wohingegen im Bereich der Treppe im linken Teil vereinzelt Personen gezählt wurden. Anders ist es in Abbildung 5b: Hier befinden sich zwei gleich große Menschenmengen bei den Tischen und Sitzmöglichkeiten sowie eine sehr groß Menge im oberen Teil des Audimax.



(a) Heatmap OG1 (b) Heatmap OG2

Abb. 5: Darstellung der gezählten Personen

Abbildung 6 zeigt den Entwurf eines Komponentendiagramms für die Anwendung (Dynamic Escape Route Guidance“ oder DERG genannt). Hierbei ist zu beachten, dass die Software unabhängig von der zugrundeliegenden Gebäudeautomation arbeiten kann. Es soll möglich sein sowohl Rauch- und Präsenzmelder über den KNX-Bus auszulesen als auch die Informationen einer Brandmeldeanlage zu verwenden. Die Daten der Personendetektion gehen über den Knoten „Human Detection“ in die Software ein und werden von dem „Input Layer“ für weitere Berechnungen vorbereitet. Diese könnten auf vorherige Analysen zurückgreifen und neue Daten in Datenbanken oder ähnlichem speichern („Persistent“ im Diagramm). Ein weiterer modularer Aufbau ist die Verwendung eines Datenmodells. In diesem werden technische Informationen über die verwendeten LED-Streifen abgebildet und der Fluchtwegplan modelliert. Diese Modularität sorgt dafür, dass die DERG an jedes Gebäude angepasst werden kann, indem lediglich das Datenmodell aktualisiert wird.

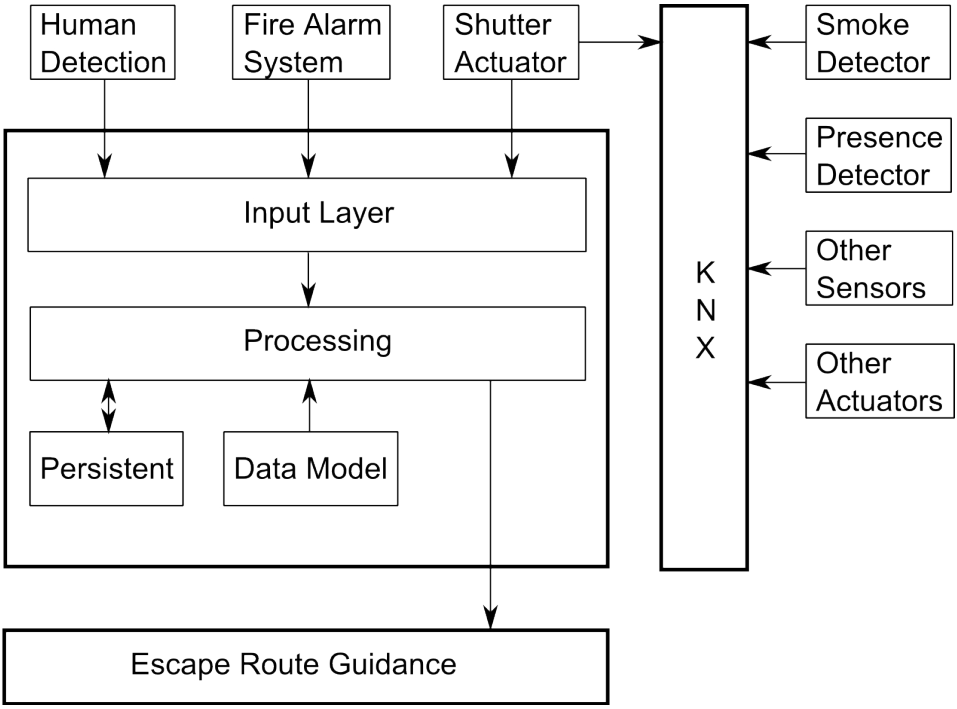


Abb. 6: Dynamic Escape Route Guidance (DERG)

Das Modellhaus ist ein erster Schritt für den modernen Brandschutz und wird Wege erforschen, um den Brandschutz mit Gebäudeautomation zu vereinen. Eine dynamische Generierung der Daten für die LED-Streifen sowie die Erkennung von Menschenmengen und die anschließende Leitung sollen dazu beitragen, dass die Personen sicherer und schneller aus einem brennenden Gebäude evakuiert werden.

Literaturverzeichnis

[Bl15] Bluetooth Basics, http://www.bluetooth.com/Pages/Basics.aspx, Stand: Mai 2015.

[Bo15] Rettungszeichen: Rettungsweg rechts (BGV A8 E 13), http://www.fluchtschilder.com/rettungszeichen/nach-BGV-A8-VGB-125/171/Rettungsweg-rechts-BGV-A8-E-13, Stand: Mai 2015.

[Co15] TrueView People Counter, http://www.cognimatics.com/Products/TrueView-People-Counter-the-leading-embedded-people-counter, Stand: Mai 2015.

[DI15] Automatische Fahrgastzählung, https://www.dilax.com/de/public-transport/automatische-fahrgastzaehlung/, Stand: Mai 2015.

[HA15a] Intelligente Leitsysteme, http://www.hanning-kahl.de/produkte-services/intelligentes-leitsystem.html, Stand: Mai 2015.

[HA15b] Vorwärts denken, http://www.hanning-kahl.de/startseite.html, Stand: Mai 2015.

[In15] Produkte — Infrarot-Messtechnik, http://www.iris-gmbh.com/infrarot-messtechnik/, Stand: Mai 2015.

[LA15] PeCo TC / TC high, http://www.peoplecounter.de/hardware/personenzaehlung/peco-tc-tc-high.html, Stand: Mai 2015.

[MG15] Personenzählung Systemübersicht, http://www.mg-industrieelektronik.de/produkte/-eyeone/proeye.php, Stand: Mai 2015.

[RF15] RFID Journal, http://www.rfidjournal.com/faq/show?139, Stand: Mai 2015.

[VI15] Kundenfrequenzanalyse, http://www.visapix.de/loesungen/maximierung-der-kundenanzahl.php, Stand: Mai 2015.

[WE15] Die NFC-Technik erklärt, http://www.pc-magazin.de/ratgeber/nfc-sicherheit-reichweite-technik-1472160.html, Stand: Mai 2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Integration of a fall detection system into the intelligent building

Aljoscha Pörtner[1], Dennis Sprute[1], Alexander Weinitschke[1] and Matthias König[1]

Abstract: Health monitoring and the integration of such systems into homely environments can support healthy aging. In this paper, we focus on the integration of a fall detection system into the intelligent building. We present a low intrusive way to detect falls and locate people inside a building based on the low power wireless technology Bluetooth Smart and a concept how well designed human computer interaction (HCI) concepts inside a building can help to save lives or at least prevent people of heavy injuries.

Keywords: fall detection, human computer interaction, smart home, intelligent building, indoor localization

1 Introduction and Motivation

The United Nations estimates that in 2050, the first time in history, our globus will accomodate more old people than young humans under 15 [UN12]. Beside the problems for our whole health and welfare system, these progress will especially make high demands on our healthcare system. To take care of these consequences and to provide a safe and viable environment for humans, preventive measures will acquire more significance. Falls constitute a large proportion of domestic accidents. They are one of the most significant causes for life-threatening injuries and show a close link between the duration until the emergency services arrive and the severity of the injury [WNI81]. Because an expansion of medical care for the growing group of elderly is not possible, measures for autonomous care and surveillance of persons will be more and more important [TV00].
This paper describes the integration of a fall detection system into an intelligent building, otherwise known as Smart Home. Beside the technical aspects of the integration, the system provides a way to integrate an information technology system into the domestic environment, and especially to achieve a better user adaption and faster initiation of the prevention in case of an accident. The system is characterized by the low costs and energy-saving components as well as good integration into the physical environment to improve the human computer interaction (HCI), which is especially important for the elderly, e.g. [Ma03].

[1] University of Applied Sciences Bielefeld, Department of Technology, 32427 Minden, Germany, [firstname.surname]@fh-bielefeld.de

2 Related Work

2.1 Integration into the intelligent building

Fall detection in general is a widely discussed field in research (see the work of Mubashir et al. [MSS13]), while the integration of those systems, especially with regard to designing HCI for elderly, receives far less attention. The acknowledged problems of smart homes in general are above all the high costs and the difficult management of the system, in combination with a lack of flexibility and inadequately developed safety concepts [Br11], which lead to a lot of obstacles especially for older people. One approach to examine and solve these problems is the project "Casa Vecchia"[5] of the Alpen-Adria University of Klagenfurt which studies ambient assisted living in rural areas interdisciplinarily by installing smart home technology into households of elderly and monitoring them scientifically. The researchers underline the calm integration of components and describe how physical objects can be used for optical signalling of variable information [Le13]. Addtionally, they conclude that a real intelligent environment should adapt to the social structure and not vice versa [LMH12]. Beside these generalized papers on technical integration and HCI inside the intelligent building, there are several specific approaches to integrate a fall detection system into the living environment. Yu et al. describe a system which detects falls visually by classification based on support vector machines (SVM) [Yu12]. The notification of a fall is realized by classic communication techniques like SMS or internet-based messaging services. In the work of Lee et al., the fall detection is supported by acceleration sensors of a standard smartphone in combination with a monitoring system and notification by email [JYK13]. Huang et al. present a system based on the wireless standard ZigBee where the fall detection is again realized by acceleration sensors. Unlike the previous systems, the position of the fallen person will be determined and sent to a Remote Telecare Provider by GSM or the Internet [HC14]. Another approach to the integration of a fall detection system is presented by Klack et al. who used a grid of piezo-electric sensors inside a sensitive floor to detect a fall and the position of the person. In the showcase, this information is visualized on a videowall [Kl11].

2.2 Wireless Networking Protocols

In the field of the wireless connections of components, a broad range of different protocols are available to choose from, where Bluetooth Smart [Bl15a], ZigBee/802.15.4 [Zi15], Z-Wave [Z-15] and ANT [AN15] seem to be future-proof and are thus often referred to in scientific literature [De13] [Si12] [Ca09]. Because power consumption is a direct quality characteristic, papers which focused on the energy efficiency of the protocols are relevant. For this topic, the papers of Demetyev et al. [De13] and Siekinnen et al. [Si12] deliver results. Both state that Bluetooth Smart is the most energy efficient variety. Demetyev et al. compared Bluetooth Smart, ZigBee/802.15.4 and ANT, while Siekkinen et al. focused

[5] Casa Vecchia, ital. = old house; project acronym: Carinthian Association of Smart Ambience – Venue Enabling Collaboration and Communication in the accustomed Home to support Independent Aging (www.casavecchia.at/).

on the comparison between Bluetooth Smart and ZigBee/802.15.4. Beside the aspect of power consumption, the focus of the technology is a quite relevant factor. Cao et al. concluded that Bluetooth Smart offers special benefits for the usage as „Wireless Body Area Networks“ [Ca09].

2.3 Indoor localization

The localization of persons or objects inside a building by radio transmission techniques is a widely discussed field in the scientific literature. Unlike the localization outside of buildings, the use of GPS is not feasable because of the high shielding of the outer building structures. The methods and techniques for indoor localization can be condensed into two approaches : fingerprinting- and model-based approaches. The fingerprinting method is based on the analysis of the environment and their surrounding signatures to build a database and uses this for a later localization process by pattern matching [YWL12]. Examples for this are RADAR [BP00] or Horus [YA05]. Model-based techniques estimate the location based on geometrical models and the behaviour of electromagnetic waves when they are transmitted [YWL12]. A lot of these models make a relationship between the signal strength and the distance between sender and receiver, e.g. the *log-distance Path loss model* (LPLM) [BS13]. Others use the Time of Arrival (ToA) [Yo06], Time Difference of Arrival (TDoA) [PCB00] or the Angle of Arrival (AoA) [NN03] to estimate the location.

3 Smart Fall System

As described in the previous section, there are many different approaches for detecting and reporting a fall. Many fall detection systems do not focus on the user-friendly and supportive integration of the system inside intelligent buildings. The system „Smart Fall“described here addresses the gap between the technical process of detecting a fall and the user involved.

3.1 System description

Our fall detection system is based on three independent components: the fall detection module, the integration module and the proper point of integration into the intelligent buildung, the base station running OpenHAB [Op15]. The fall detection module is a small wearable device, while the integration module is a single-board computer. Communication between these submodules is realized by both the wireless standard Bluetooth Smart (former known as Bluetooth Low Energy) and wireless LAN. Figure 1 shows the associated structure of the system. Here we see the monitored person equipped with the fall detection module. Additionally, there are transponders which are located in every room near the door to make the process of localization possible. Every transponder is integrated in the network

by wireless LAN. For complete coverage in wider areas, additional transponders can be installed in any location inside the building. The base station, in form of an OpenHAB, can be positioned at an arbitrary location with the restriction that it must be connected with the network by LAN or WLAN. Beside the described submodules, LED light strips are embedded into the floor. These strips are connected to the system by KNX [KN15] which is a standardized, OSI-based network communications protocol for intelligent buildings. The process of communication between the module *fall detection and localization* and the KNX module *LED light strip* is realized by the OpenHAB and its integrated „bindings“.

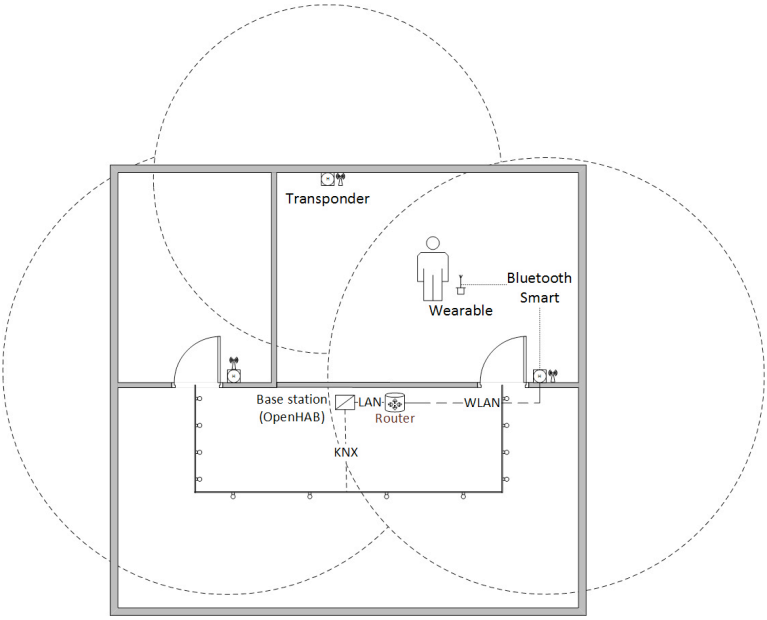


Fig. 1: System architecture Smart Fall

3.2 Scenario description

The environment of the scenario is an assisted living complex for elderly. In this special form of housing, the elderly live mostly independently with a high standard of self determination and with concurrent support by trained staff as well as neighbors. Person A, in the following Alice, and person B, in the following Bob, are equipped with a fall detection module. The whole facility is prepared for the operation of such system. This includes a regular deployment of the components and the installation of light strips. In this example every apartment consists of three rooms. The rooms are identified by the label pattern „apartment.roomnumber“(Figure 2). In our example Alice is located in room *A.1*. To use the radio, she walks into room *A.2*. When she walks through the door, the check-in process in room *A.2* is detected by the transponder and reported to the base station. Now Alice stumbled over the edge of the carpet and falls, bumping her head and entering a state of quantitative consciousness disturbance. Now she is not able to move by herself or to get help. The fall detection module detects this state and reports a fall to the base station,

which processes and aggregates both, the previously obtained localization data and the fall data. As a result of the emergency, the emergency notification process will be initiated, which consists of two subprocesses. On the one hand, professional help will be called via the notification of a carer or doctor and on the other hand, for the first aid, a neighbor will be informed. Enabled through the process of localization, the system is capable of accomplishing the localization of Bob which is located in room *B.3*. The fall detection system draws attention to the emergency and supports Bob with the discovery of the accident patient by activating the lights strips which guide him to Alice.

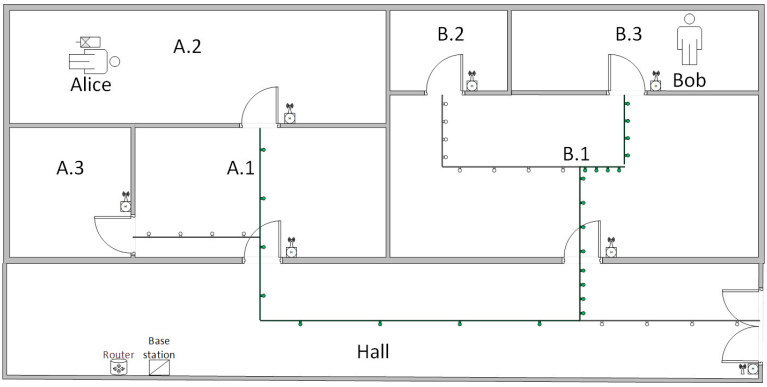


Fig. 2: Scenario description „Fall“

3.3 Hardware

The following section deals with the components used by the system and the underlying hardware. As described in brief in Chapter 3.1, the system consists of three different physical components. On the user side, a wearable consisting of two main parts is located: A System on Chip (SoC), which is made up of a 32-bit ARM Cortex M0 with 256 KB flash memory, 16 KB memory and a Bluetooth Smart component (Table 1) and a tri-axial acceleration sensor measuring linear acceleration in three orthogonal directions. The component is designed as an ultra low power Integrated Circuit (IC) and shows a power consumption of 130 μA, which leads to a greatly extended battery life. The accelerometer supports a bandwidth up to 1kHz with a resolution of 12 bit and a range of $\pm$ 16g (Table 2).

CPU	ARM Cortex M0, 32-bit
Flash	256 kB
RAM	16 kB
Digital IO	SPI, I2C, UART
Power supply	1.8 V - 3.6 V

Tab. 1: SoC key characteristics

Measurement range	$\pm$2, $\pm$4, $\pm$8, $\pm$16 g
Digital resolution	12 bit
Bandwidth	1000 Hz - 8 Hz
Digital IO	SPI, I2C
Power supply	2.4 V - 3.6 V

Tab. 2: Accelerometer key characteristics

CPU	Allwinner A10 (Cortex A8)s, 32-bit
Weight	14 g
RAM	512 MB DDR3-SDRAM
Digital IO	USB-2.0-Host, Micro-HDMI (Typ D), Mini-USB, GPIO FPC
Power supply	5V/150mA - 5V/220mA

Tab. 3: Integration module key characteristics

Both components are assembled on a 2.6 cm x 3 cm sized chip (Figure 3). The second essential component is the transponder or integration module. Due to its small size and the high performance and low power consumption, we use a single-board computer. It consists of a Allwinner A10s with a Cortex A8, 512 MB DDR3-SDRAM and a power consumption of 150 up to 220 mA. The Bluetooth Smart connection is realized by a Bluetooth 4.0 enabled USB-Dongle. The total dimension of the transponder is 40 x 40 x 32 mm (Figure 4).

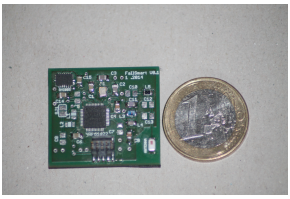


Fig. 3: Fall detection module

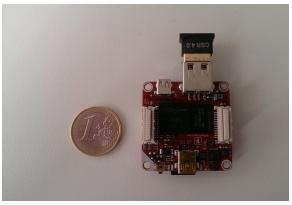


Fig. 4: Integration module

3.4 Fall detection

We describe the fall detection sensor briefly. The fall detection itself is realized by a threshold-based method (TBM). The acceleration data are collected by the tri-axial sensor and evaluated on the SoC. The foundation for the process of evaluation is a state machine with the five states : *Before Fall*, *Free Fall*, *Impact*, *After Impact* and *Fall*. Before a fall occurs, the measured acceleration value is nearly 1 g, which is caused by the gravitational acceleration and its force. In the transition to a fall, there is a short period of time with an acceleration value close to 0 g, before the impact causes a value $> 3g$. In order to be able to make the distinction between a real fall and a similar situation, e.g. stair climbing, the user's posture is analyzed and classified as „Stable "or „Unstable " based on the horizontal or vertical pose of the person after the impact. A detailed description of the approach can be found in [Sp15].

3.5 Integration into the intelligent building

A lot of the described systems in Chapter 2 have no or only rudimentary integration into intelligent buildings, which is mostly realized by messaging services, e.g. SMS or E-Mail. The following system presents an approach to integrating a fall detection system into the intelligent building with particular regard to user adoption and the involvement of the physical environment.

3.5.1 Bluetooth Smart

To enable the integration of the system into the intelligent building, the system is founded upon established standards in the building automation. Based on the insights discussed in Chapter 2 regarding wireless network standards and their power consumption with respect to their operationality in WBANs, we have chosen Bluetooth Smart as the wireless connection standard.
Bluetooth Smart, formerly known as Bluetooth Low Energy, was announced in 2010 as part of the new Bluetooth standard 4 and provides advantages particularly with regard to energy efficiency and interoperability. The basis of the technology is a hierarchical profile definition (Generic Attribute Profile) with services and characteristics contained therein. In case of the Smart Fall system, a service with four characteristics is defined, which enables the fall detection component to provide values, e.g. dimensional acceleration data, for the transponders. Further, the characteristics *VSA* (vector sum of acceleration data) as well as *Fall* which is a binary flag for the fall indication are assigned with a notification property which enables the providing component to notify the receiving components only in the case of a change in value (Table 4). This behaviour enables a power saving process in which every component can stay in low energy mode until a possible fall is detected and reported.

Name	UUID	Format	Properties
XVal	00001411-0000-1000-8000-00805f9b34fb	IEEE-754 32-bit Float	READ
YVal	00001412-0000-1000-8000-00805f9b34fb	IEEE-754 32-bit Float	READ
YVal	00001413-0000-1000-8000-00805f9b34fb	IEEE-754 32-bit Float	READ
VSA	00001414-0000-1000-8000-00805f9b34fb	IEEE-754 32-bit Float	READ, NOTIFY
Fall	00001415-0000-1000-8000-00805f9b34fb	Boolean	READ, NOTIFY

Tab. 4: Smart Fall Service Definition

3.5.2 OpenHAB

The set of problems related to the heterogeneous protocol landscape in wireless networks, introduced in Chapter 3.5.1, can be transferred on the wired connections as well. The most popular representatives are KNX, HomeMatic [Ho15] and iNELS [iN15]. To address these problems, OpenHAB was developed. The system is a vendor- and protocol independent integration platform with low hardware requirements and a userfriendly App- and Web-based user interface to provide an easy way to configure and control home automation components. With an integrated domain specific language, OpenHAB provides a low-level approach to program procedures and behaviour in the context of home automation. The event-based architecture enables the integration of all kinds of different devices and components which are addressed over specific protocol bindings. For the integration of the Smart Fall system into the OpenHAB ecosystem, a proprietary binding was implemented which provides the interface to the various transponders and enables the localization and the notification of falls.

3.5.3 Localization

ITU indoor propagation model The localization or the distance estimation is realized with a model-based approach founded on the *ITU indoor propagation model* described in [In12]. The model is based on the path loss prediction which is an empirical mathematical formulation for the characterization of radio wave propagation. The general *ITU model* is a site-generic model founded on the insights of the free-space loss model [BKW11] (Equation 1) derived by the Friis power transmission equation [Fr46]

$$FSPL(dB) = 10log_{10}\left(\left(\frac{\lambda}{4\pi d}\right)^2\right), \tag{1}$$

where λ is the wavelength in meter and d is the transmitter to receiver distance in meter. The free space path loss model is only applicable in an area with no RF obstacles. Over time, researchers have applied a path loss exponent n which is intended to stabilize the model in different environments. Based on this, the ITU released the *ITU model for indoor attenuation* (Equation 2) which handles the particular obstacles in buildings, e.g walls, and estimates the path loss an indoor link may experience.

$$L_{total} = 20log_{10}(f) + N * log_{10}(d) + Lf(n) - 28 \quad [dB] \tag{2}$$

In Equation 2 is f the frequency of the protocol in Mhz, which is approximately 2400 MHz for Bluetooth Smart [Bl15b]. N is a coefficient for the power loss of the connection which is specified in [In12]. The variable $Lf(n)$ describes the floor penetration factor which is determined as $4n$ in residential apartments where n is the number of floors [In12].
For our purpose, N and $Lf(n)$ are not sufficient for Bluetooth Smart. Our empirical evaluation of the values determined $N = 16$ and $Lf(n) = 14$ as parameters for an acceptable modeling of the path loss, especially for lower distances (Figure 5). The distance between sender and receiver is represented as d. With the knowledge of the total pathloss $L_{total} = P_{Tx_{dBm}} - P_{Rx_{dBm}}$ between sender and receiver, we can estimate the distance d and use this for the localization process. Because the Bluetooth RSSI data is subjected to electrical interference, an estimation based on the RSSI raw data is nearly impossible [JKB13]. For this reason, we use the arithmetic mean to stabilize the measured data and make a coarsely estimation possible (Equation 3).

$$A = \frac{1}{n}\sum_{i=1}^{n} L_{i_{total}} \tag{3}$$

We determined a value $n = 30$ as an appropriate value. As the setpoint value for the check-in process, we set a limit of $d < 1m$. In case of this value being undercut, the transponder notifies the base station which then decides if it is a check-in or a check-out for the corresponding person in a specific room.

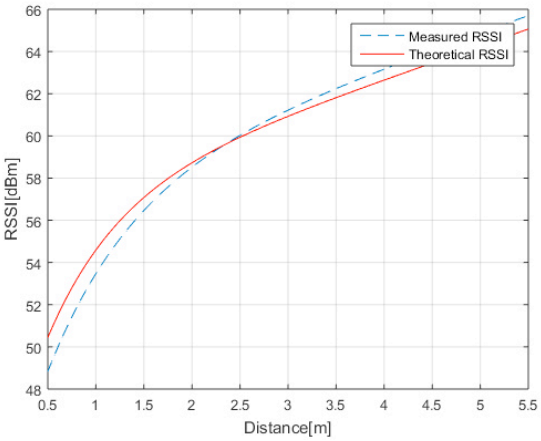


Fig. 5: ITU Model RSSI vs. Measured RSSI

3.5.4 Human computer interaction

This chapter discusses the concrete integration of the process into the physical homely environment of the elderly. We use an LED light strip actuator based on the KNX network communication protocol for intelligent buildings. As briefly described in Chapter 3.2, after a fall the base station will be informed and the alert process begins. First, the fallen person will be located based on the previously collected data. Then, a potential rescuer will be determined on an individual basis. The next step is the inclusion of the physical environment. We use the KNX-binding of OpenHAB to control the light strip by sending data telegrams which are converted into actions. Every LED in the light strip can be swiched on or off to visualize a path on the floor from any room A to any other room B. Moreover, a chaser light can be toggled which visualizes the direction of the path and enables other persons to find the right room immediatly (Figure 7).



Fig. 6: LED light strip off



Fig. 7: LED light strip on

4 Conclusion and Future Work

We presented the principles of our system to detect a fall and locate a person, discussed the concrete integration into the home environment of the elderly. According to the HCI

framework of Saizmaa et al. [SK08], our interaction inside the system focuses especially on the Human/Technology (H/T) and Technology/Home (T/H) perspective. On the one hand the ease of use and acceptability is a very important point but the automation and standardised characteristics are necessary as well. To address these points, we have designed a system which uses a very minimally invasive method of fall detection. Other systems are founded on camera-based approaches or use a lot of large hardware. Through the usage of an non-proprietary integration system without any cloud architecture like OpenHAB, we have made a high level of automation possible without the loss of privacy and self determination. A further important point is the ubiquitous characteristic of the system. The whole system consists of small sized components, which are easy to deploy. All of the processes are designed to make a very low need of adaption for the user possible. The report functionality of the system is embedded into the physical environment of a home which leads in the end to a very minimally intrusive technology with focus on the user and not on the technology itself.
In future work, we will consider the possibilities to integrate more building components, e.g. smart locks or smart windows, to make the acute care possible faster and more efficiently. Further, we want to determine how we can use sensors like air quality, temperature or sound intensity to enable the system to report a brief overview of the environmental situation to the skilled personnel and to stabilize and improve the situation for the accident victim.

References

[AN15] ANT Wireless: , ANT Basics. Website, 2015. Retrieved from `http://www.thisisant.com/developer/ant/ant-basics/` (2015, April 20).

[BKW11] Burbank, Jack L.; Kasch, William; Ward, Jon: An introduction to network modeling and simulation for the practicing engineer. ComSoc guides to communications technologies. John Wiley & Sons, 2011.

[Bl15a] Bluetooth SIG: , Bluetooth Core Specification 4.2. Website, 2015. Retrieved from `https://www.bluetooth.org/en-us/specification/adopted-specifications\?_ga=1.235534465.1491899422.1429453009` (2015, April 20).

[Bl15b] Bluetooth SIG Inc.: , Radio Architecture - Frequency Bands and Channel Arrangement. Website, 2015. Retrieved from `https://developer.bluetooth.org/TechnologyOverview/Pages/Radio.aspx` (2015, April 20).

[BP00] Bahl, Paramvir; Padmanabhan, Venkata N: RADAR: An in-building RF-based user location and tracking system. In: INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE. volume 2. Ieee, pp. 775–784, 2000.

[Br11] Brush, AJ; Lee, Bongshin; Mahajan, Ratul; Agarwal, Sharad; Saroiu, Stefan; Dixon, Colin: Home automation in the wild: challenges and opportunities. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 2115–2124, 2011.

[BS13] Botta, Miroslav; Simek, Milan: Adaptive distance estimation based on RSSI in 802.15. 4 network. Radioengineering, 22(4), 2013.

[Ca09] Cao, Huasong; Leung, Victor; Chow, Cupid; Chan, Henry: Enabling technologies for wireless body area networks: A survey and outlook. Communications Magazine, IEEE, 47(12):84–93, 2009.

[De13] Dementyev, Artem; Hodges, Steve; Taylor, Stuart; Smith, Joshua: Power consumption analysis of Bluetooth Low Energy, ZigBee and ANT sensor nodes in a cyclic sleep scenario. In: Wireless Symposium (IWS), 2013 IEEE International. IEEE, pp. 1–4, 2013.

[Fr46] Friis, H.T.: A Note on a Simple Transmission Formula. Proceedings of the IRE, 34(5):254–256, May 1946.

[HC14] Huang, Chih-Ning; Chan, Chia-Tai: A ZigBee-based location-aware fall detection system for improving elderly telecare. International journal of environmental research and public health, 11(4):4233–4248, 2014.

[Ho15] HomeMatic: , HomeMatic – wir machen Home Control einfach. Website, 2015. Retrieved from `http://www.homematic.com/`(2015, April 20).

[In12] International Telecommunication Union: Recommendation ITU-R P.1238-7 - Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 900 MHz to 100 GHz. Technical report, International Telecommunication Union, 2012.

[iN15] iNELS: , iNELS BUS System. Website, 2015. Retrieved from `http://www.inels.com/products/inels-bus-system`(2015, April 20).

[JKB13] Jung, Joonyoung; Kang, Dongoh; Bae, Changseok: Distance Estimation of Smart Device using Bluetooth. In: ICSNC 2013, The Eighth International Conference on Systems and Networks Communications. pp. 13–18, 2013.

[JYK13] Jer-Vui Lee; Yea-Dat Chuah; Kenny T.H. Chieng: Smart Elderly Home Monitoring System with an Android Phone. In (Science and Engineering Research Support Society, ed.): International Journal of Smart Home, volume Vol.7 No.3 2013, pp. 17–32. SERSC, 2013.

[Kl11] Klack, Lars; Möllering, Christian; Ziefle, Martina; Schmitz-Rode, Thomas: Future Care Floor: A Sensitive Floor for Movement Monitoring and Fall Detection in Home Environments. In (Akan, Ozgur; Bellavista, Paolo; et al., eds): Wireless Mobile Communication and Healthcare, volume 55 of Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pp. 211–218. Springer Berlin Heidelberg, 2011.

[KN15] KNX: , What is KNX? Website, 2015. Retrieved from `http://www.knx.org/knx-en/knx/association/what-is-knx/index.php` (2015, April 20).

[Le13] Leitner, Gerhard; Hitz, Martin; Fercher, Dipl-Ing Anton J; Brown, John NA: Aspekte der Human Computer Interaction im Smart Home. HMD Praxis der Wirtschaftsinformatik, 50(6):37–47, 2013.

[LMH12] Leitner, Gerhard; Melcher, Rudolf; Hitz, Martin: Spielregeln im intelligenten Wohnumfeld. In: Vernetzung als soziales und technisches Paradigma, pp. 189–206. Springer, 2012.

[Ma03] Marcus, Aaron: Universal, Ubiquitous, User-interface Design for the Disabled and Elderly. ACM Interactions, 10(2):23–27, March 2003.

[MSS13] Mubashir, Muhammad; Shao, Ling; Seed, Luke: A survey on fall detection: Principles and approaches. Neurocomputing, 100:144–152, 2013.

[NN03] Niculescu, D.; Nath, B.: Ad hoc positioning system (APS) using AOA. In: INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies. volume 3, pp. 1734–1743 vol.3, March 2003.

[Op15] OpenHAB: , OpenHAB - empowering the smart home. Website, 2015. Retrieved from `http://www.openhab.org/` (2015, April 20).

[PCB00] Priyantha, Nissanka B.; Chakraborty, Anit; Balakrishnan, Hari: The Cricket Location-support System. In: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking. MobiCom '00, ACM, New York, NY, USA, pp. 32–43, 2000.

[Si12] Siekkinen, M.; Hiienkari, M.; Nurminen, J.K.; Nieminen, J.: How low energy is bluetooth low energy? Comparative measurements with ZigBee/802.15.4. In: Wireless Communications and Networking Conference Workshops (WCNCW), 2012 IEEE. pp. 232–237, April 2012.

[SK08] Saizmaa, T.; Kim, Hee-Cheol: A Holistic Understanding of HCI Perspectives on Smart Home. In: Networked Computing and Advanced Information Management, 2008. NCM '08. Fourth International Conference on. volume 2, pp. 59–65, Sept 2008.

[Sp15] Sprute, Dennis; Pörtner, Aljoscha; Weinitschke, Alexander; König, Matthias: Smart Fall: Accelerometer-Based Fall Detection in a Smart Home Environment. In: Inclusive Smart Cities and e-Health, volume 9102 of Lecture Notes in Computer Science, pp. 194–205. Springer International Publishing, 2015.

[TV00] Tang, Puay; Venables, Tim: 'Smart'homes and telecare for independent living. Journal of Telemedicine and Telecare, 6(1):8–14, 2000.

[UN12] UNFPA and HelpAge International: Ageing in the Twenty-First Century. Report, UNFPA and HelpAge International, 2012.

[WNI81] Wild, Deidre; Nayak, US; Isaacs, B: How dangerous are falls in old people at home? BMJ, 282(6260):266–268, 1981.

[YA05] Youssef, Moustafa; Agrawala, Ashok: The Horus WLAN location determination system. In: Proceedings of the 3rd international conference on Mobile systems, applications, and services. ACM, pp. 205–218, 2005.

[Yo06] Youssef, Moustafa; Youssef, Adel; Rieger, Chuck; Shankar, Udaya; Agrawala, Ashok: Pin-Point: An Asynchronous Time-Based Location Determination System. In: Proceedings of the 4th International Conference on Mobile Systems, Applications and Services. MobiSys '06, ACM, New York, NY, USA, pp. 165–176, 2006.

[Yu12] Yu, Miao; Rhuma, A.; Naqvi, S.M.; Wang, Liang; Chambers, J.: A Posture Recognition-Based Fall Detection System for Monitoring an Elderly Person in a Smart Home Environment. Information Technology in Biomedicine, IEEE Transactions on, 16(6):1274–1286, Nov 2012.

[YWL12] Yang, Zheng; Wu, Chenshu; Liu, Yunhao: Locating in Fingerprint Space: Wireless Indoor Localization with Little Human Intervention. In: Proceedings of the 18th Annual International Conference on Mobile Computing and Networking. Mobicom '12, ACM, New York, NY, USA, pp. 269–280, 2012.

[Z-15] Z-Wave Alliance: , About Z-Wave Technology. Website, 2015. Retrieved from `http://z-wavealliance.org/about_z-wave_technology/` (2015, April 20).

[Zi15] ZigBee Alliance: , ZigBee: The Open, Global Wireless Standard for Connecting Everyday Devices. Website, 2015. Retrieved from `http://www.zigbee.org/zigbee-for-developers/` (2015, April 20).

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Die Wirkung von Feedback und Goal Setting auf den Energieverbrauch von Privathaushalten – Eine Meta-Analytische Untersuchung

Tobias Weiß[1], Madlen Diesing[1] und Marco Krause[1]

Abstract: Der Energieverbrauch von Privatkunden ist bisweilen intransparent. Mit dem Ausbau intelligenter Gebäudetechnologien in Smart Homes geht auch der verstärkte Einsatz von Feedbacksystemen einher. Diese sollen die Transparenz des Verbrauchs erhöhen und somit die Einsparung wertvoller Ressourcen unterstützen. Im Rahmen der Untersuchung erfolgt eine Analyse und Systematisierung von Studien und Pilotprojekten rund um die Themen Feedback bzw. Goal-Setting sowie eine Beurteilung der Auswirkungen auf den Energieverbrauch. Das Ziel ist die Schaffung einer Grundlage zum Entwurf eines Entscheidungsunterstützungssystems. Als Ergebnis trat hervor, dass vergangenheitsbezogene Vergleiche und hohe Übermittlungsfrequenzen des Feedbacks besonders relevante Elemente sind. Soziale Vergleiche hingegen haben nur eine begrenzte Wirkung. Gerätespezifische Angaben werden von Nutzern als sehr sinnvoll wahrgenommen, ebenso die Angabe der Verbrauchswerte in Geldeinheiten. Dabei treten zahlreiche Forschungslücken zutage, welchen in nachfolgenden Untersuchungen begegnet werden.

Keywords: feedback, goal setting, decision support, energy intelligence, business intelligence

1 Einleitung

Der Energiewandel ist eines der wichtigsten Themen in den Medien. Diese Diskussionen gehen einher mit vielfältigen technischen Unterstützungsmöglichkeiten, welche zudem einen deutlichen Wandel der Interaktion zwischen Technologie und Nutzer bewirken. In dem Kontext der intelligenten Gebäudetechnologie werden oft Begriffe wie bspw. Smart Meter genannt. Aktuell fanden Untersuchungen mit Smart Metern, zumindest in Deutschland, hauptsächlich in Pilotprojekten statt. Auch aus anderen Ländern sind zahlreiche Informationen bekannt, ebenso wie Datenmaterial von bereits ausgerollten Projekten. Ein Teil dieser Untersuchungen hinterfragt dabei den Mehrwert, der aus den neu entstehenden Datenmengen generiert werden kann. Diese Daten können gewinnbringend ausgewertet werden wie Potentialuntersuchungen aufzeigen. Aber haben die eigentlichen Endkunden der Energie, die Privathaushalte, auch einen realen Nutzen von dieser erhöhten Technisierung? Es eröffnet sich an dieser Stelle ein Spannungsfeld zwischen dem Nutzen der Energieversorger und dem der Konsumenten, welche aus der Interaktion mit neuen intelligenten Gebäudetechnologien hervortreten.

[1] Technische Universität Dresden, Lehrstuhl für Wirtschaftsinformatik – Business Intelligence Research, Münchner Platz 3, 01187 Dresden, tobias.weiss@tu-dresden.de

1.1 Motivation

Der eingehenden Literaturmeinung folgend, würde dem Nutzer ein Mehrwert durch die Aufbereitung dieser Daten entstehen (vgl. [SSR12], S. 89). Dies erfolgt durch eine Erhöhung der Transparenz des Energieverbrauchs - ein deutlicher Fortschritt zur aktuell bekannten, jährlichen Abrechnung, bei der der Kontext zwischen Ursache und (finanzieller) Auswirkung aufgrund der zeitlichen Diskrepanz verloren geht. Ein zeitnahes Identifizieren und Intervenieren bezüglich kostenintensiver Tätigkeiten wird so durch die verstärkte Interaktion zwischen dem Nutzer und der intelligenten Gebäudetechnologie realisierbar. Weiterhin haben Energieversorgungsunternehmen zu Zeiten eines hohen Wettbewerbs und niedriger Margen ein verstärktes Interesse daran, abseits des Wettbewerbs über einen möglichst niedrigen Preis, Bestandskunden zu halten und Neukunden durch Zusatzangebote zu begeistern (vgl. [We14]). Die Bereitstellung von aktuellen Verbrauchsdaten hat daher unterschiedliche Motivationsgrundlagen und wird in zahlreichen (Pilotprojekt-)Studien umfassend dargestellt.

Im Rahmen einer Voruntersuchung wurde die Unübersichtlichkeit des Forschungsfeldes deutlich. So beschreiben die Studien jeweils nur einen Teilbereich an Darstellungsformen für Feedback und Goal-Setting, in vielen Fällen ohne Hinterfragen der tatsächlichen Eignung (ist bspw. die Darstellung des Verbrauchswertes der letzten Woche in kWh für den Privatkunden verständlich?). Ist der Konsument nicht in der Lage, aus den gebotenen Optimierungsmöglichkeiten eigene Vorteile zu generieren, rekurriert dieser Tatbestand negativ auf die Motivation zur Nutzung solcher intelligenten Technologien. Bereits existente Metastudien behandeln zumeist einen anderen Fokus. Es lassen sich zwar Ableitungen von grundlegenden Darstellungselementen für Entscheidungsunterstützungssysteme vornehmen, diese sind jedoch nicht zugeschnitten auf die Branchendomäne mit speziellen Anforderungen (z.B. geringes Involvement der Anwender aufgrund des Charakters des Produkts „Strom“ als unemotionales, immer identisches Commodity-Produkt). Zum Aufbau eines möglichst optimalen Dashboards für Privatkunden im Rahmen der Entscheidungsunterstützung ist jedoch eine Zerlegung der Darstellung in einzelne Komponenten sowie eine Bewertung der Wirksamkeit hinsichtlich des Einsparziels erforderlich.

2 Methodisches Vorgehen

2.1 Forschungsziel

Das Forschungsziel ist die Entwicklung einer Systematisierung aus Komponenten zur Unterstützung des Entwurfs von Feedback und Goal Setting im Rahmen eines Entscheidungsunterstützungssystems zur Reduktion des Stromverbrauchs in Privathaushalten. Dabei sollen, im Rahmen eines Erkenntnisziels, vorhandene Gestaltungsmöglichkeiten von Feedback und Goal Setting identifiziert und deren Wirkung auf den Stromverbrauch und die Präferenzen der Nutzer erhoben werden. Im

Rahmen des Gestaltungsziels erfolgt dann eine Systematisierung dieser Gestaltungsmöglichkeiten. Folgende **Forschungsfragen** stehen dabei im Fokus:

1. Welche Elemente existieren zur Gestaltung von Feedback und Goal Setting und können in der Interaktion zwischen dem Nutzer und der intelligenten Technologie einen wirksamen Beitrag zur Reduktion des Stromverbrauchs in Privathaushalten leisten?
2. Welche Präferenzen haben Privatanwender bzgl. identifizierter Elemente?

2.2 Forschungsmethode

Unter der wissenschaftstheoretischen Position des gemäßigten Konstruktivismus durch theoriebasierte Exploration wird die Methode der Literaturanalyse angewendet. Diese ist nach Bonfadelli & Meier als ein Typ der Metaanalyse zu verstehen (vgl. [BM84], S. 539), welche dem Ziel der Synthese und Generalisierung mit analytischem Bezug zu vorhandenem Datenmaterial folgt (vgl. [Co98], S. 106; [Hs88], S. 311 ff.). Der Kritik an der Literaturanalyse („nicht systematische und kaum standardisierte Vorgehensweise“ ([BM84], S. 542)) wird durch ein strukturiertes Vorgehen nach Fettke Rechnung getragen, orientiert an Cooper (vgl. [Fe06], [CH94]). Es handelt sich um eine integrative Literaturrecherche mit dem Ziel der Darstellung des Forschungsstandes und Systematisierung (vgl. [CH94], S. 3). Die Datenextraktion erfolgt dabei in Einzelarbeit, wurde jedoch durch einen weiteren Experten zur Qualitätssicherung begutachtet.

3 Grundlagen

3.1 Einflussfaktoren auf den Stromverbrauch

Die Gründe für schwankenden Energieverbrauch in Haushalten sind komplex und variieren stark. Anhand der Literatur lassen sich Einflussfaktoren in Gruppen einteilen:

Soziodemografische Faktoren: können nicht variiert werden, tragen aber zur Erklärung der Verbräuche bei (Wohnfläche (vgl. [MC79], S. 170), Haustyp (vgl. [GAL13], S. 737), Bewohneranzahl (vgl. [MC79], S. 170), Alter (vgl. [Mo05], S. 271), Einkommen (vgl. [GAL13], S. 737) und Bildung (vgl. [GAL13], S. 738))

Energiepreise (vgl. [GAL13], S. 737)

Psychologische Faktoren: Beeinflussbar durch Feedback und Goal Setting (Interesse und Einstellung (vgl. [LST13], S. 1315), Motivation (intrinsisch (vgl. [KW12], S. 51) / extrinsisch (vgl. [SBD14], S.104), kulturelle und soziale Normen (vgl. [LST13], S. 1315), Wissen (vgl. [Kr13]) und Bewusstsein (vgl. [CH13], S. 9))

Die Reduktion des Stromverbrauchs im Haushalt lässt sich durch eine Änderung des gewohnheitsmäßigen oder kaufbezogenen Verhaltens bewirken. Unter **gewohnheitsmäßiges Verhalten** versteht man alltägliche Aktivitäten im Haushalt, z.B. das Ausschalten des Lichts in ungenutzten Räumen. **Kaufbezogenes Verhalten** bezieht sich bspw. auf den Neuerwerb eines effizienteren Kühlgerätes. Diese Handlungen erfordern finanzielle und technische Ressourcen, sind daher mittel- bis langfristiger Natur (vgl. [BGF05], S. 1425f). Beide Komponenten können über die Instrumente Feedback und Goal Setting angesprochen werden.

3.2 Feedback und Goal Setting

Feedback definiert sich als eine Handlung, bei der jemandem von einer externen Quelle Informationen über die Erfüllung einer bestimmten Aufgabe bereitgestellt werden (vgl. [KD96], S. 255). Im konkreten Fall werden den Haushalten vom Energieversorger bzw. Gebäudeinformationssystem Informationen über den eigenen Stromverbrauch bereitgestellt.

Ziele hingegen stellen einen Bezugspunkt dar, der zur Bewertung der Informationen dient, die durch Feedback zur Verfügung gestellt werden. Daraus geht hervor, dass **Goal Setting** nicht allein, sondern immer in Kombination mit Feedback angewendet werden muss. Nur so kann der Nutzer erkennen, in welche Richtung er sich relativ zum Ziel entwickelt (vgl. [MM02], S. 591). Dabei wird unterschieden in implizite und explizite Ziele (vgl. [SBD14], S.100). Die Systematisierung der Studien bezieht sich ausschließlich auf die Wirksamkeit von expliziten Zielen, bei denen ein konkreter Wert der Einsparung vorgegeben wird. Explizite Ziele sind im Rahmen der Zielerreichung des Performance Managements direkt messbar, was bei impliziten Zielen, die vom Nutzer determiniert und ohne Dokumentation kontrolliert werden, nicht der Fall ist (vgl. [SBD14], S. 104).

Die vorgestellten Komponenten Feedback sowie Goal Setting sollen dabei folgende **Funktionen** erfüllen: Aufmerksamkeit erregen (vgl. [GT11], S. 142), Wissen erhöhen (vgl. [HV89], S. 99; [GS96], S. 83), Zielerreichung dokumentieren (vgl. [HV89], S. 99; [GS96], S. 83) und Einstellungen beeinflussen (vgl. [HV89], S. 99; [GS96], S. 88).

4 Vorgehen der systematischen Literaturrecherche

4.1 Planung

Folgende Suchbegriffe wurden aus dem Forschungsziel abgeleitet:

Tab. 1: Suchbegriffe (eigene Darstellung)

Kategorie	Thema	Zielgruppe	Ziel I	Ziel II	Instrumente
	Energie	Private Haushalte	Verringerung des Stromverbrauchs	Nutzerpräferenzen	Feedback Goal Setting
Begriffe	*energy*	household*	consum*	prefer*	feedback
		residential*	saving*	favor*	"goal setting"
		domestic*	conserv*		"goal-setting"
		home*	reduc*		

Die Kategorien wurden mit AND verknüpft, die Suchbegriffe innerhalb einer Kategorie mit OR, nach folgendem Schema: *Thema AND Zielgruppe AND Ziel I / II AND Instrument => Feedback / Goal Setting*. Die Kombinationen wurden in den Datenbanken *Sciencedirect* sowie *EBSCO Host Academic Search Complete* angewendet. Weiterhin wurden die Literaturverzeichnisse der aktuellsten Studien (Ende Jahr 2013) aufgrund der aktuellen Entwicklung per Rückwärtssuche ausgewertet.

4.2 Ausschluss- und Qualitätskriterien

Zur Filterung der Suchergebnisse wurden folgende Ausschluss- und Qualitätskriterien aus den Forschungsfragen entwickelt:

Ausschlusskriterien:

Die Studie untersucht weder die Wirkung einer Intervention auf den Energieverbrauch noch Nutzerpräferenzen.

Die untersuchte Zielgruppe der Studie sind **keine privaten Haushalte** (z.B. Verhalten am Arbeitsplatz oftmals gänzlich verschieden (vgl. [Ab05], S. 274)).

Das Ziel der Studie ist die **Verschiebung des Energieverbrauchs** auf andere Tageszeiten (Lastverlagerung ist weder zielführend zur Reduktion des Gesamtverbrauchs (vgl. [Fi08], S. 87) noch vorteilhaft für den Haushalt (vgl. [SSR12], S. 210)).

Die betrachteten Instrumente sind **nicht Feedback und/oder Goal Setting**.

Die Studie ist **keine Primärstudie** und sie ist nicht im **Volltext** verfügbar.

Die Studie ist **nicht in den Sprachen Deutsch oder Englisch** verfasst.

Die Studie untersucht **nicht die Energieart Strom** (Haushalte reagieren bei unterschiedlichen Energiearten verschieden (vgl. [Hu86], [Mi83])).

Qualitätskriterien:

Teilnahme an der Studie: freiwillig vs. nicht freiwillig (bei ausschließlich freiwilliger Teilnahme ohne zufallsbasierten Stichprobenverfahren der Haushalte besteht die Gefahr, dass die Ergebnisse nicht repräsentativ sind (vgl. [Da13], S. 402))

Aufteilung der Teilnehmer auf verschiedene Untersuchungsgruppen: zufällig vs. nicht zufällig (nicht zufällige Zuteilungen können zu strukturellen Unterschieden führen, die Effekte auf das Ergebnis haben können (vgl. [Da13], S. 402))

Abgangsrate (als Abgangsrate wird die Zahl der Haushalte betrachtet, die nicht bei der Auswertung der Studie berücksichtigt wurden (vgl. [Da13], S. 403))

Vorhandensein **statistischer Analyse** (Prüfen der Ergebnisse auf Signifikanz um einen Zufall auszuschließen (vgl. [Ki04], S. 15))

Nutzung einer **Kontrollgruppe** (Ausschluss von Zeitdauereffekte (vgl. [Hü05], S. 161))

4.3 Datenextraktion und Datensynthese

In Anlehnung an [Ki04] wurden folgende Daten extrahiert und dokumentiert:

Quantitative Studien:

Eigenschaften: Autor, Jahr, Name, Land, Art der Studie, Aufbau, Energieart

Rahmenbedingungen: Dauer, Eigenschaften der Teilnehmer, Zahlungsmodalitäten, Teilnehmerzahl

Eigenschaften Instrumente: Medium, Übermittlungsfrequenz, Eigenschaften Feedback, Eigenschaften Goal Setting

Ergebnisse und Qualität: Einsparungen, Signifikanz, Qualitätskriterien 1-5

Präferenzstudien (Untersuchung von Nutzerpräferenzen)

Eigenschaften: Autor, Jahr, Name, Land, Untersuchungsmethode, Energieart

Rahmenbedingungen: Eigenschaften der Teilnehmer, Teilnehmerzahl

Im Rahmen der Datensynthese werden die Ergebnisse aller Studien auf qualitativem Weg mittels einer Kategorisierung der Gestaltungsmöglichkeiten von Feedback und Goal Setting zusammengeführt. Mithilfe der quantitativen Studien erfolgt im nächsten Schritt eine Bewertung der Eignung hinsichtlich des Ziels der Reduktion des Stromverbrauchs, sowie eine Integration der Hauptaussagen der Präferenzstudien.

5 Auswertung der systematischen Literaturrecherche

Die initiale Suche ergab 252 Ergebnisse. Nach einer Bereinigung von Duplikaten verblieben 191 Studien, die anhand der Ausschlusskriterien (siehe Kapitel 4.2) auf Eignung geprüft wurden. Nach der Filterung verblieben 23 Studien. Die Auswertung der Literaturverzeichnisse ergab 35 weitere, relevante Studien. Es ergab sich final eine Anzahl von **58 Studien**, die der Inhaltsanalyse zugeführt wurden.

Von den 58 Studien messen **38 Studien quantitativ** die Auswirkungen von Feedback und Goal Setting. Im Rahmen der in Kapitel 4.3 beschriebenen Datenextraktionsschritte wurden alle verfügbaren Informationen entnommen und dokumentiert. **21 Studien** befragen die Nutzer nach Ihren Wünschen und Vorlieben und stellen damit die in Kapitel 4.3 benannten Präferenzstudien dar. Eine Studie untersucht beide Themengebiete und ist demnach doppelt aufgeführt.

Die Qualität der quantitativen Experimente variiert stark. Als wichtigste Qualitätskriterien wurden die Kriterien 4 (Vorhandensein statistischer Analyse) und 5 (Kontrollgruppe) festgelegt. Studien, die diese Kriterien nicht erfüllen, wurden von weiteren Betrachtungen ausgeschlossen. Dies hat zum Ausschluss von fünf Studien geführt. Von den 33 Studien untersuchen nur 16 systematisch die unterschiedlichen Wirkungen einzelner Gestaltungsmöglichkeiten. Systematisch bedeutet hier, dass ein Element unter sonst gleichen Bedingungen variiert wurde.

Die Daten der verbliebenen Studien wurden extrahiert und die Ergebnisse für jedes in Kapitel 4.3 aufgeführte Merkmal übersichtlich in einer Tabelle dokumentiert. Aufgrund der Reichhaltigkeit der Informationen wird an dieser Stelle auf eine detaillierte Darstellung verzichtet, da im Fokus der Publikation die Feedback- und Goal-Setting-Elemente stehen.

6 Datensynthese: Die Elemente von Feedback und Goal Setting

6.1 Systematische Übersicht

Wie in der vorliegenden Studie bereits erwähnt, ist eine Anwendung von Goal Setting ohne Feedback nicht sinnvoll. Dies wurde durch die Studie von Becker (vgl. [Be78]) bestätigt. Demnach findet eine Integration von Goal Setting in den Gesamtkontext von Feedback statt. Die Gesamtheit wird fortan als Feedbacksystem bezeichnet. Die nachfolgende Abbildung zeigt eine systematische Übersicht der identifizierten Elemente, wobei eine Untergliederung in obligatorische und optionale Elemente vorgenommen wurde. Obligatorische Elemente müssen in jedem Feedbacksystem auftauchen, sind jedoch von den gewählten Ausprägungen der optionalen Elemente abhängig.

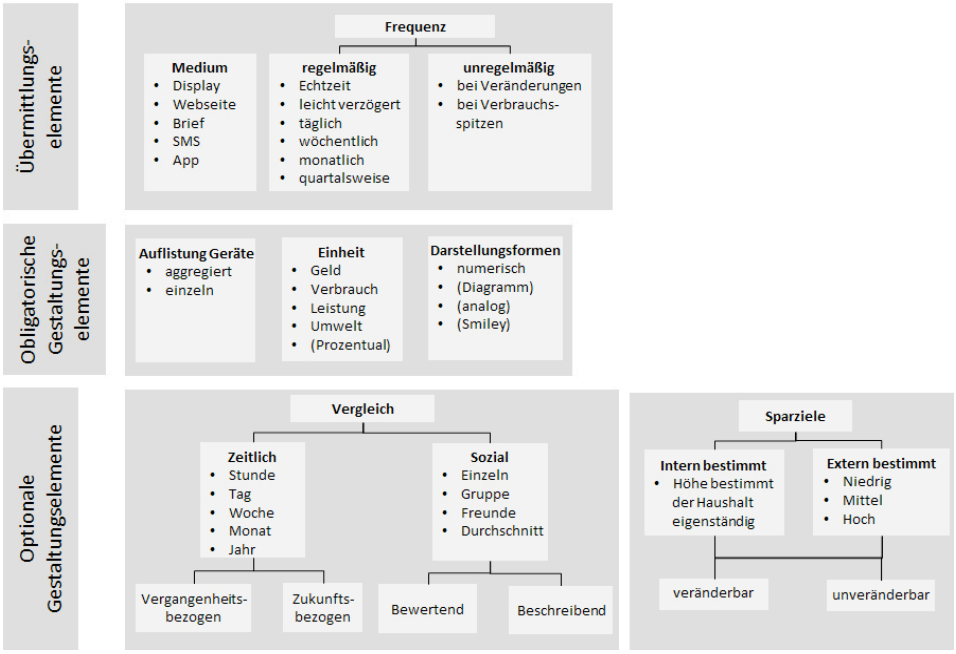


Abb. 1: Systematische Übersicht der Elemente eines Feedbacksystems (eigene Darstellung)

6.2 Erläuterungen zu den Kategorien und Elementen

Tabelle 2 beschreibt die Aussagen der zur strukturierten Analyse herangezogenen Literaturquellen. Dabei erfolgt eine Aufteilung in die drei Schichten der Elemente des Feedbacksystems: optionale, obligatorische Gestaltungselemente sowie Übermittlungselemente.

Tab. 2: Beurteilung der optionalen Gestaltungselemente des Feedbacksystems

Optionale Gestaltungselemente		**Beurteilung**
Zeitliche Vergleiche	Zukunfts-bezogen	• Hypothese: verbrauchsmindernde Wirkung (vgl. [GAL13]) • Keine systematischen Untersuchungen vorhanden, Forschungsbedarf erforderlich

	Vergan-genheits-bezogen	• verbrauchsmindernde Wirkung (vgl. [Fi08], S. 101) • Nutzerpräferenzen stark ausgeprägt (vgl. [BDR12], S. 392; [Ka11], S. 466; [AW09], S. 3; [PSW09], S. 2851; [RHH04], S. 4) • möglichst detaillierte Vergleiche (vgl. [GAL13])
Soziale Vergleiche	Beschrei-bend	• Keine Auswirkung, positive Effekte gleichen sich mit negativen Effekten aus (Animation von Geringverbrauchern zur Steigerung des Verbrauchs durch Vergleich) (vgl. [Sc07], S. 430; [Fi08], S. 101; [SLL11]) • verbrauchsmindernde Wirkung bei zusätzlicher Integration von bewertenden und beschreibenden Elementen (zur Beibehaltung einer positiven Tendenz) (vgl. [Sc07], [SLL11], [Al11]) • Nutzerentscheidung über Verwendung afgr. unter-schiedlicher Präferenzen (Zustimmung: [FS09], S. 41, Ablehnung: [RHH04], S. 4, Allg. Skepsis: [FS09], S. 41, [RHH04], S. 14, [Pe11], S. 28)
	Bewertend	
Ziele	Extern vor-gegeben	• Goal Setting als essentielles Element eines Feedbacksystems (vgl. [MM02]) • Höhe des Ziels hat einen Einfluss auf die Wirksamkeit: verbrauchsmindernde Wirkung bei Vorgabe von 15-20% (vgl. [MM02], [LST13], [Be78]) • verbrauchsmindernde Wirkung bei Selbstvorgabe (vgl. [LST13]) • Nutzerentscheidung über Verwendung, aufgrund unterschiedlicher Präferenzen (Hilfreich: [BDR12], S. 393; [AW09], S. 3; Ablehnung: [FS09], S. 42)
	Selbst gewählt	

Tab. 3: Beurteilung der obligatorischen Gestaltungselemente des Feedbacksystems

Obligatorische Gestaltungs-elemente	**Beurteilung**

Auflistung elektronischer Geräte	• Forderungen in den Studien sowie Nutzerpräferenzen bezüglich gerätespezifischer Auflistung vorhanden (vgl. [Fi08]; [BDR12], S. 392; [VWD12], S. 99; [We12], S. 662; [Ka11], S. 466) • aggregierte Anzeige afgr. Komplexität jedoch ausreichend (vgl. [FS09], S. 41; [Pi10], S. 247) und durch Studien belegt (vgl. [GAL13], [Al11]) • Forschungsbedarf erforderlich (aggregierte Darstellung vs. gerätespezifische Darstellung)
Einheiten	• Nutzerpräferenz: Darstellung in Geldeinheiten (vgl. [Kr13], S. 457; [Ka11], S. 466; [AW09], S. 3; [FS09], S. 40) • Geldeinheiten mit hohem Detailgrad vermeiden (Wirkung der Geringhaftigkeit der Einsparungen) (vgl. [FS09], S. 40; [WOV11], S. 151; [Ka11], S. 466) • Verbrauchs (kWh)- und Umwelteinheiten (CO_2) ergänzend (vgl. [AW09], S. 3; [FS09], S. 40) • Forschungsbedarf bzgl. Wirkung der Einheiten auf das Verständnis und die Einsparungen erforderlich
Darstellungs-formen	• detaillierte numerische Werte (keine Geldeinheiten), ergänzt um Smileys oder analoge Darstellung ([Al11]; [BDR12], S. 393; [FS09], S. 42; [GAL13]) • Geschlechterspezifische Unterschiede (Männer: Tendenz zu konkreten Daten, Frauen: Ampelfarben) (vgl. [St08], S. 14) • Einfachheit der Darstellung (vgl. [HNB10], S. 6118) • Keine konkreten Aussagen zu Darstellungsformen oder allgemeinen Konsens: Forschungsbedarf

Tab. 4: Beurteilung der Übermittlungselemente des Feedbacksystems (eigene Darstellung)

Übermittlungs-elemente		**Beurteilung**
Fre-qu-enz	regel-mäßig	• besonders starke verbrauchsmindernde Wirkung bei hoher Frequenz (vgl. [Fi08], S. 97; [Ab05], S. 281; [GAL13]; [Al11])
	unregel mäßig	• ergänzende Übermittlung bei Verbrauchsspitzen kann hilfreich sein (vgl. [FS09], S. 40) • Kombination einer wählbaren Frequenz mit Benachrichtigungsfunktionen (vgl. [Gl10])

Medium	• elektronische Medien um hohe Übermittlungsfrequenz abzubilden (vgl. [Fi08]) • Medium muss präsent sein (z.B. Display, vgl. [Ke11], S. 70) • Webseite: Gefahr des „Vergessen-Werdens“ (vgl. [Ke11], S. 70), dennoch beliebt (vgl. [FS09], S. 43; [BDR12], S. 393)

7 Fazit

Es ist zunächst anzumerken, dass Goal Setting als ein Element des Feedbacksystems identifiziert wurde. Dieses sowie alle weiteren Elemente und deren Ausprägungen wurden im Kapitel 6 erläutert. Die sich ergebende Systematisierung ist in Abbildung 1 ersichtlich.

Elemente, die einen wirksamen Beitrag zur Reduktion des Stromverbrauchs in Privathaushalten leisten können, wurden mit den tabellarischen Zusammenfassungen der Auswertungen im Rahmen der Forschungsarbeit in Kapitel 6.2 aufgezeigt. Besonders hervorzuheben sind dabei detaillierte vergangenheitsbezogene Vergleiche sowie eine hohe Übermittlungsfrequenz über elektronische Medien. Soziale Vergleiche sowie Ziele können ebenfalls verbrauchsreduzierende Wirkung haben, sind aber mit Hinblick auf die Nutzerpräferenzen mit Vorsicht zu betrachten. Während der Beurteilung wurde auf die Eignung im Zusammenhang mit anderen Elementen hingewiesen. Dies betrifft bspw. die detaillierte numerische Darstellung, die jedoch nicht in Geldeinheiten erfolgen sollte. Weiterhin sind im sozialen Vergleich beschreibende und bewertende Elemente erforderlich. Auch der Zusammenhang zwischen Übermittlungsfrequenz und -medium wurde herausgestellt.

Die Präferenzen von Privatanwendern liegen bei vergangenheitsbezogenen Vergleichen, gerätespezifischen Angaben, Anzeigen in Geldeinheiten sowie hohen Übermittlungsfrequenzen. Bei sozialen Vergleichen, Zielen und Übermittlungsmedien ist jedoch auf die unterschiedlichen Vorlieben der Nutzer hinzuweisen.

Bei der Systematisierung und Beurteilung der Elemente ist stellenweise ein starker Forschungsbedarf aufgefallen, obwohl bereits zahlreiche Studien die Auswirkungen von Feedbacksystemen auf den Energieverbrauch untersuchen. Diese Forschungslücken sollen in nachfolgenden Studien eingehender untersucht werden, wobei einige Fragestellungen nur durch tatsächliche, quantitative Pilotprojekte zu untersuchen sind. Weiteres Ausbaupotential liegt in der Betrachtung weiterer Energiearten. So können z.B. zusätzliche Einsparungen im Gasverbrauch erreicht werden. Denn nur eine ganzheitliche Effizienzsteigerung bei der Nutzung aller Energieressourcen kann bei der Reduktion des Kohlenstoffdioxidausstoßes zielführend sein und bringt einen volkswirtschaftlichen Mehrwert.

Mit der vorliegenden Arbeit konnte dennoch ein wichtiger Schritt zur Systematisierung

des teilweise unübersichtlichen und in jeweils unterschiedlichen Facetten betrachtenden Literaturbestandes getan werden. Diese Analyse dient somit als Grundstein zur Konstruktion und Evaluierung eines prototypischen Entscheidungsunterstützungssystems für Privathaushalte, welches die Präferenzen und Kompetenzen der Anwender möglichst gut mit den nutzenversprechenden Elementen eines Feedbacksystems verbinden soll. Dabei gilt es noch anzumerken, dass die an die Kompetenzen des Verbrauchers angepasste Darstellung der Verbrauchsoptimierung, sowie der in Aussicht gestellte Nutzen, hierbei als Schlüsselfaktoren für ein Feedbacksystem fungieren. Diese akzentuieren die resultierenden Vorteile für den Endverbraucher und dadurch die intelligente Technologie in Gebäuden, sie legitimeren damit die verbundene Technik-Nutzer-Interaktion aus Verbrauchersicht.

Literaturverzeichnis

[Ab05] Abrahamse, W.; Steg, L.; Vlek, C.; Rothengatter, T.: A review of intervention studies aimed at household energy conservation. In: Journal of Environmental Psychology 25 (3), S. 273–291, 2005.

[Al11] Allcott, H.: Social norms and energy conservation. In: Journal of Public Economics 95 (9-10), S. 1082–1095, 2011.

[AW09] Anderson, W.; White, V.: Exploring consumer preferences for home energy display functionality. Online verfügbar unter http://ecoeye.co.uk/academic/CSE-Report.pdf, zuletzt geprüft am 14.11.2013, 2009.

[BDR12] Bonino, D.; Corno, F.; Russis, L.: Home energy consumption feedback: A user survey. In: Energy and Buildings 47, S. 383–393, 2012.

[Be78] Becker, L. J.: Joint effect of Feedback and Goal Setting on Performance. In: Journal of Applied Psychology 63 (4), S. 428–433, 1978.

[BGF05] Barr, S.; Gilg, A. W.; Ford, N.: The household energy gap. In: Energy Policy 33 (11), S. 1425–1444, 2005.

[BM84] Bonfadelli, H.; Meier, W.: Meta-Forschung in der Publizistikwissenschaft. In: Rundfunk und Fernsehen 32. Jg. H. 4, S. 537-550, 1984.

[Ch13] Chiang, T.; Mevlevioglu, G.; Natarajan, S.; Padget, J.; Walker, I.: Inducing [sub]conscious energy behaviour through visually displayed energy information. Accepted Manuscript. In: Energy and Buildings, 2013.

[CH94] Cooper, H.; Hedges, L. V.: Research Synthesis As a Scientific Enterprise. In: Harris Cooper; Larry V. Hedges (Hrsg.): The Handbook of Research Synthesis. New York, S. 3–14, 1994.

[Co98] Cooper, H.: Synthesizing Research. A Guide for Literature Reviews. 3. Aufl. Thousand Oaks, London, 1998.

[Da13] Davis, A..; Krishnamurti, T.; Fischhoff, B.; Bruine de Bruin, W.: Setting a standard for electricity pilot studies. In: Energy Policy 62, S. 401–409, 2013.

[DBH10] van Dam, S. S.; Bakker, C. A.; van Hal, J. D. M.: Home energy monitors: impact over the medium-term. In: Building Research & Information 38 (5), S. 458–469, 2010.

[Fe06] Fettke, P.: Eine Untersuchung der Forschungsmethode "Review" innerhalb der Wirtschaftsinformatik. In: Wirtschaftsinformatik (4), S. 257–266, 2006.

[Fi08] Fischer, C.: Feedback on household electricity consumption: a tool for saving energy? In: Energy Efficiency 1 (1), S. 79–104, 2008.

[FS09] Fitzpatrick, G.; Smith, G.: Technology-Enabled Feedback on Domestic Energy Consumption. In: IEEE Pervasive Comput. 8 (1), S. 37–44, 2009.

[GAL13] Gans, W.; Alberini, A.; Longo, A.: Smart meter devices and the effect of feedback on residential electricity consumption. In: Energy Economics 36, S. 729–743, 2013.

[Gl10] Gleerup, M.; Larsen, A.; Leth-Petersen, S.; Togeby, M.: The effect of feedback by Text Message (SMS) and Email on Household Electricity Consumption. In: Energy Journal 31 (3), S. 113–132, 2010.

[GS96] Gardner, G. T.; Stern, P. C.: Environmental problems and human behavior. Boston: Allyn and Bacon, 1996.

[GT11] Grønhøj, A.; Thøgersen, J.: Feedback on household electricity consumption. In: International Journal of Consumer Studies 35 (2), S. 138–145, 2011.

[HNB10] Hargreaves, T.; Nye, M.; Burgess, J.: Making energy visible. In: Special Section: Transition Pathways to a Low Carbon Economy 38 (10), S. 6111–6119, 2010.

[Hs88] Hsia, H. J.: Mass Communications Research Methods: A Step-by-Step Approach. Hillsdale, New Jersey, Hove, London: Lawrence Erlbaum Associates, 1988.

[Hü05] Hübler, O.: Einführung in die empirische Wirtschaftsforschung. Probleme, Methoden und Anwendungen. München: Oldenbourg, 2005.

[Hu86] Hutton, R. B.; Mauser, G. A.; Filiatrault, P.; Ahtola, O. T.: Effects of Cost-Related Feedback on Consumer Knowledge and Consumption Behavior. In: Journal of Consumer Research (13), S. 327–336, 1986.

[HV89] van Houwelingen, J. H.; Van Raaij, W. F.: The Effect of Goal-Setting and Daily Electronic Feedback on In-Home Energy Use. In: Journal of Consumer Research. Jun89, Vol. 16 Issue 1, 16 (1), S. 98–105, 1989.

[Ka11] Karjalainen, S.: Consumer preferences for feedback on household electricity consumption. In: Energy & Buildings 43 (2/3), S. 458–467, 2011.

[KD96] Kluger, A. N.; DeNisi, A.: The effects of feedback interventions on performance. In: Psychological Bulletin 119 (2), S. 254–284, 1996.

[Ke11] Kerrigan, D. J.; Gamberini, L.; Spagnolli, A.; Jacucci, G.: Smart meters: A users' view. In: PsychNology Journal 9 (1), S. 55–72, 2011.

[Ki04] Kitchenham, B.: Procedures for Performing Systematic Reviews. Keele University, zuletzt geprüft am 13.12.2013, 2004.

[Kr13] Krishnamurti, T.; Davis, A. L.; Wong-Parodi, G.; Wang, J.; Canfield, C.: Creating an in-home display. In: Applied Energy 108, S. 448–458, 2013.

[KW12] Kua, H. W.; Wong, S. E.: Lessons for integrated household energy conservation policies from an intervention study in Singapore. In: Energy Policy 47, S. 49–56, 2012.

[LST13] Loock, C.-M.; Staake, T.; Thiesse, F.: Motivating Energy-Efficient Behavior with Green IS. In: MIS Quarterly 37 (4), S. 1313–1322, 2013.

[MC79] McClelland, L.; Cook, S. W.: Energy Conservation Effects of Continuous In-Home Feedback in All-Electric Homes. In: Journal of Environmental Systems 9 (2), S. 169–173, 1979.

[Mi83] Midden, C. J. H.; Meter, J. F.; Weenig, M. H.; Zieverink, H. J. A.: Using feedback, reinforcement and information to reduce energy consumption in households. In: Journal of Economic Psychology 3 (1), S. 65–86, 1983.

[MM02] McCalley, L.T; Midden, C. J.H.: Energy conservation through product-integrated feedback. In: Journal of Economic Psychology 23 (5), S. 589–603, 2002.

[Mo05] Moll, H. C.; Noorman, K. J.; Kok, R.; Engström, R.; Throne-Holst, H.; Clark, C.: Pursuing More Sustainable Consumption by Analyzing Household Metabolism in European Countries and Cities. In: Journal of Industrial Ecology 9 (1-2), S. 259–275, 2005.

[Pe11] Petkov, P.; Köbler, F.; Foth, M.; Krcmar, H.: Motivating domestic energy conservation through comparative, community-based feedback in mobile and social media. In: Marcus Foth, Jesper Kjeldskov und Jeni Paay (Hg.): the 5th International Conference on Communities and Technologies. Brisbane, Australia, 02.07.2011, S. 21–30, 2011.

[Pi10] Pierce, J.; Fan, C.; Lomas, D.; Marcu, G.; Paulos, E.: Some consideration on the (in)effectiveness of residential energy feedback systems. In: Olav W. Bertelsen und Peter Krogh (Hg.): the 8th ACM Conference on Designing Interactive Systems. Aarhus, Denmark, S. 244–247, 2010.

[PSW09] Petersen, D.; Steele, J.; Wilkerson, J.: WattBot: A residential Electricity Monitoring and Feedback System. In: Dan R. Olsen, Richard B. Arthur, Ken Hinckley, Meredith Ringel Morris, Scott Hudson und Saul Greenberg (Hg.): the 27th international conference extended abstracts of human factors in Computing Systems. Boston, MA, USA, S. 2847–2852, 2009.

[RHH04] Roberts, S.; Humphries, H.; Hyldon, V.: Consumer Preferences for Improving Energy Consumption Feedback. Centre for Sustainable Energy. Online verfügbar unter https://www.ofgem.gov.uk/ofgem-publications/58008/8144-consumerfdbakpref.pdf, zuletzt geprüft am 15.11.2013, 2004.

[SBD14] Seligman, C.; Becker, L. J.; Darley, J. M.: Encouraging Residential Energy Conservation Through feedback. In: A. Baum, J. E. Singer und Jerome L. Singer (Hg.): Advances in Environmental Psychology. Volume 3: Energy Conservation, Psychological Perspectives. Hoboken: Taylor and Francis, S. 93–113, 2014.

[Sc07] Schultz, P. W.; Nolan, J. M.; Cialdini, R. B.; Goldstein, N. J.; Griskevicius, V.: The Constructive, Destructive, and Reconstructive Power of Social Norms. In: Psychological Science 18 (5), S. 429–434, 2007.

[SLL11] Staake, Thorsten; Landwehr, Jan; Loock, Claire-Michelle: Green IS Design and Energy Conservation: An Empirical Investigation of Social Normative Feedback. In:

32nd International Conference on Information Systems. Shanghai, S. 1–15. Online verfügbar unter http://aisel.aisnet.org/icis2011/proceedings/humanbehavior/10, 2011.

[SSR12] Servatius, H.-G.; Schneidewind, U.; Rohlfing, D.: Smart Energy. Wandel zu einem nachhaltigen Energiesystem. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2012.

[St08] Strengers, Y.: Smart metering demand management programs: challanging the comfort and cleanliness habitus of households. In: Nic Bidwell (Hg.): the 20th Australasian Conference on Computer-Human Interaction: Designung for Habitus and Habitat. Cairns, Australia, S. 9–16, 2008.

[VWD12] Vassileva, I.; Wallin, F.; Dahlquist, E.: Understanding energy consumption behavior for future demand response strategy development. In: Energy 46 (1), S. 94–100, 2012.

[We12] Weiss, M.; Staake, T.; Mattern, F.; Fleisch, E.: PowerPedia: changing energy usage with the help of a community-based smartphone application. In: Personal & Ubiquitous Computing 16 (6), S. 655–664, 2012.

[We14] Weiß, T: Potentiale von Business Intelligence für innovative Zusatzdienste mit Daten aus dem Smart Metering. In: 13. Symposium Energieinnovation, Graz, S. 296-297, 2014.

[WOV11] Wallenborn, G.; Orsini, M.; Vanhaverbeke, J.: Household appropriation of electricity monitors. In: International Journal of Consumer Studies 35 (2), S. 146–152, 2011.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Betriebliche Umweltinformationssysteme (BUIS)

Environmental Management Information Systems (EMIS)

17. Tagung der Fachgruppe Betriebliche Umweltinformationssysteme der Gesellschaft für Informatik (GI) e.V.
7. BUIS-Tage

Workshop: Betriebliche Umweltinformationssysteme (7. BUIS-Tage 2015)

Corinna V. Lang[1], Jorge Marx Gómez[2], Volker Wohlgemuth[3]

Abstract: Betriebliche Bemühungen um Umweltschonung, Nachhaltigkeit und Energieeffizienz werden durch den Einsatz von Umweltmanagementsystemen unterstützt, die eine Vielzahl heterogener Daten benötigen, um der Planung, Steuerung und Kontrolle von Umweltaufgaben gerecht zu werden. Diese Daten müssen durch Betriebliche Umweltinformationssysteme (BUIS) vorgehalten und zielorientiert bearbeitet werden. Dieser Workshop, der von der Fachgruppe BUIS zum 17. Mal durchgeführt wird, zum 7. Mal untern dem Titel „BUIS-Tage“, informiert über den aktuellen Stand von BUIS und stellt neue Ideen und Lösungsansätze vor, die in einer Expertenrunde aus Wissenschaft und Praxis intensiv erörtert werden.

Keywords: Fachgruppe Betriebliche Umweltinformationssysteme, Energiemanagement, Integrated Management Systems, Wassermanagement, Nachhaltigkeitsmanagement, BUIS

Den Mitgliedern des Programmkomitees danken wir sehr für ihre Expertise bei der Ausgestaltung des Workshop-Programms sowie bei der fachlichen Begutachtung der eingereichten Beiträge.

1 Programmkomitee

- Burkhardt Funk, Leuphana Universität Lüneburg
- Corinna V. Lang, Hochschule Anhalt (Chair)
- Jorge Marx Gómez, Universität Oldenburg (Chair)
- Peter Niemeyer, Leuphana Universität Lüneburg
- Volker Wohlgemuth, Hochschule für Technik und Wirtschaft Berlin (Chair)

2 Beiträge

Es wurden 9 Beiträge angenommen, die nach eingehender Begutachtung als für den Workshop geeignet bewertet wurden. Die Themen reichen von Anwendungsbeispielen zum Einsatz von BUIS über Integrationsfragestellungen bis hin zu Unterstützung des ökologischen Produktdesigns.

[1] Hochschule Anhalt, c.lang@wi.hs-anhalt.de
[2] Universität Oldenburg, jorge.marx.gomez@uni-oldenburg.de
[3] Hochschule für Technik und Wirtschaft Berlin, volker.wohlgemuth@htw-berlin.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Ein Überblick über Einsatzbereiche von betrieblichen Umweltinformationssystemen (BUIS) in der Praxis

Volker Wohlgemuth[1]

Abstract: Der nachfolgende Beitrag gibt einen Überblick über die Einsatzbereiche von betrieblichen Umweltinformationssystemen (BUIS) in der Praxis und leitet hieraus sich ergebene Entwicklungspfade ab. Es wird hier vornehmlich der Status Quo der gegenwärtigen Softwarelandschaft im Bereich der Produktion und Planung betrieblicher Maßnahmen im Kontext des betrieblichen Umweltschutzes dargestellt. Zur besseren Strukturierung wurden ausgehend von den drei Hauptaufgabenbereichen der BUIS Untergruppen gebildet, mittels derer sich die vielen verschiedenen ausgewählten Vertreter der BUIS besser zuordnen ließen. Hieraus werden Fragen in Bezug auf die zukünftige Entwicklung von BUIS abgeleitet.

Keywords: Betriebliche Umweltinformationssysteme, BUIS, Umweltmanagement, Digitale Fabrik.

1 Einleitung und Motivation

Seit Anfang der 90er Jahre des letzten Jahrhunderts beschäftigt sich die betriebliche Umweltinformatik mit sog. betrieblichen Umweltinformationssystemen (BUIS) [HR95]. BUIS dienen ganz allgemein dazu, umweltrelevante Daten und Informationen eines Betriebes zu erfassen, zu dokumentieren und danach so aufzubereiten, dass diese für unterschiedliche Problemstellungen des betrieblichen Umweltschutzes genutzt werden können, z.B. für die Dokumentation und Sicherstellung behördlicher Auflagen und Genehmigungen, zur Simulation von Stoffströmen in der Produktion oder der Planung, Steuerung und Kontrolle betrieblicher Umweltschutzmaßnahmen [Wo05]. In diesem Sinne stellen BUIS eine spezielle Ausprägung klassischer Umweltinformationssysteme (sog. UIS) dar, die bereits auf eine längere Historie zurück blicken können und überwiegend bei Umweltbehörden und Forschungsinstituten für Aufgaben des Umweltschutzes, der Umweltplanung und Umweltforschung eingesetzt werden [HR95]. BUIS stehen jedoch aufgrund ihres Fokus auf Betriebe und der Tatsache, dass ein signifikanter Teil umweltrelevanter Informationen schon in anderen Informationssystemen vorliegt, in engen Zusammenhang mit den traditionellen betrieblichen Informationssystemen. Daher sind BUIS als Erkenntnisobjekt im Überlappungsbereich von Wirtschafts- und Umweltinformatik anzusiedeln [HR95]. Es besteht somit eine gewisse Analogie zwischen betrieblichen Informationssystemen und BUIS, was Möglichkeiten bietet, betriebswirtschaftliche und umweltbezogene Ziele bei der Gestaltung betrieblicher Informationssysteme zu kombinieren. In diesem Sinne ist anzumerken, dass das Thema BUIS

[1] HTW Berlin, Fachbereich 2, Wilhelminenhofstraße 75A, 12459 Berlin, volker.wohlgemuth@htw-berlin.de

in der Gesellschaft für Informatik (GI) e.V. von den Fachbereichen „Informatik in den Lebenswissenschaften“ und „Wirtschaftsinformatik“ gemeinsam initiiert wurde und bis heute kontinuierlich in der Fachgruppe Betriebliche Umweltinformationssysteme weiterentwickelt wird [HR95].

Der Begriff des BUIS wird in der Fachliteratur sehr unterschiedlich und vielfältig benutzt und steht dabei je nach Fachrichtung nicht für identische Inhalte [Wo05]. Eine Übersicht über Definitionen von BUIS findet sich z.B. bei [Ju10]. In diesem Beitrag wird in Anlehnung an [Be02] und [HW09] bewusst eine stark an der Praxis des betrieblichen Umweltschutzes ausgerichtete Definition verwendet, um den real existierenden betrieblichen Einsatzfeldern von BUIS gerecht zu werden. Unter einem betrieblichen Umweltinformationssystem (BUIS) wird hier demnach jede Softwareanwendung verstanden, die für die Erfassung, Dokumentation und Bewertung betrieblicher Umweltwirkungen sowie zur Generierung, Planung und Steuerung vom Umweltschutzmaßnahmen genutzt wird und damit das betriebliche Umwelt- und Nachhaltigkeitsmanagement in seinen Aufgaben zumindest in gewissen Teilaspekten unterstützt. Diese Definition ist relativ weit gefasst, spiegelt jedoch die betriebliche Praxis der Anwendung von Softwarewerkzeugen wider, die auf dem Gebiet des betrieblichen Umweltschutzes und Nachhaltigkeitsmanagements verwendet werden. Als Softwaresysteme müssen BUIS die relevanten Umweltdaten in Form von Stoff- und Materialstammdaten (Eigenschaft der Stoffe), Strukturdaten (Organisationsstruktur und Prozessabfolge), Prozessdaten (funktionelle Zusammenhänge), Stoff- und Energieflussdaten (Art und Menge) und Organisationsdaten (organisatorische Abläufe) verarbeiten können [La03]. Hierzu müssen diese Daten administriert, ausgewertet, visualisiert und für Modellrechnungen und Simulationen verwendet werden können [HR95]. Bei der Integration in die Systemlandschaft des Unternehmens können BUIS hierbei umgesetzt sein als [HR95]:

- Stand-Alone-Systeme (Insellösung),
- Add-on-System (über Schnittstellen integriert in andere Informationssysteme) oder
- Integraler Bestandteil vorhandener Systeme

Die Praxis sowie bisherige wissenschaftliche Analysen zeigen jedoch, dass BUIS zumeist hauptsächlich als Stand-Alone-Systeme umgesetzt werden [JL11].

Dieser Beitrag soll hauptsächlich einen Einblick in die Praxis der Einsatzvielfalt von BUIS im Sinne obiger Definition geben. Allerdings sind hier Tabellenkalkulationsanwendungen herausgenommen worden, da diese in der Regel über nur eine geringe Komplexität verfügen und nicht fachspezifisch ausgelegt sind. Es sollen danach mögliche Entwicklungspfade für eine zukünftige Entwicklung von BUIS aufgezeigt werden, die sich allerdings nur aus den Anforderungen der Praxis ergeben. So wurden für die Analyse die Tagungsbände der Fachgruppe Betriebliche Umweltinformationssysteme der GI herangezogen sowie vom Autor sowohl aus der Automobilindustrie als

auch mit kleinen und mittleren Unternehmen (KMU) gesammelte Erfahrungen eingearbeitet. Entwicklungstendenzen, die sich aus der Wissenschaft und Forschungsaktivitäten der betrieblichen Umweltinformatik ergeben, werden in diesem Beitrag bewusst ausgeklammert, um den Fokus nicht zu verwässern.

2 Grundlegende Aufgabenbereiche von BUIS

Der Einsatzbereich von BUIS in der betrieblichen Praxis ist vielfältig und vielgestaltig. Nachfolgend wird der Versuch unternommen, diese unterschiedlichsten BUIS gemäß ihrer Aufgabenbereiche im betrieblichen Umweltschutz zu klassifizieren. Im Kern lassen sich drei Hauptaufgabenbereiche ableiten [Wo05], [Hi95]:

1. Informationsbeschaffung und Dokumentation
2. Bewertung und Entscheidungsunterstützung
3. Steuerung, Planung und Kontrolle

Auf Basis dieser drei Hauptgruppen wurde eine praxisorientierte Sammlung von beispielhaften BUIS erstellt und in dahinter stehende Untergruppen, wie sie in der nachfolgenden Abbildung dargestellt sind, identifiziert. Dadurch soll versucht werden, die Softwarewerkzeuge im Bereich der BUIS allgemein besser zu klassifizieren.

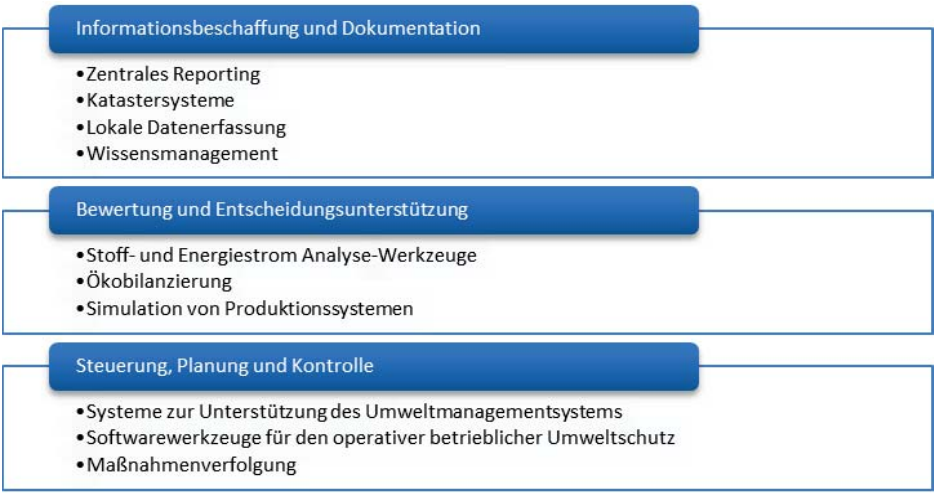


Abb. 1: Aufgabenbereiche von BUIS

Jedoch enthalten machen BUIS Funktionen, die eine eindeutige Einordung in diese Gruppen nicht immer ganz eindeutig zulassen. So dienen Stoffstromanalysewerkzeuge

auch der Informationsbeschaffung, indem diese die IST-Situation der stofflichen und energetischen Seite der betrieblichen Auswirkungen auf die Umwelt darstellen. Ferner sollen auf einer Analyse und Bewertung dieser Stoffströme Maßnahmen zur Verbesserung und Optimierung abgeleitet werden und dienen somit hauptsächlich einer Entscheidungsunterstützung. Auch bieten einige Softwarewerkzeuge Lösungen für verschiedene der genannten Aufgabenbereiche, indem sich beispielsweise Module oder Bausteine für mehrere Aufgabenbereiche hinzufügen lassen (wie z.B. bei dem BUIS EcoWebDesk[2]). Aus diesem Grunde lassen sich manche BUIS nicht eindeutig einem der genannten Aufgabenbereich zuordnen, sondern bedienen mehrere Anforderungen.

3 Informationsbeschaffung und Dokumentation

Eine wichtige Aufgabe eines BUIS ist die Ermittlung und Darstellung der betrieblichen Umweltauswirkungen und hieraus abgeleiteter umweltrelevanter Informationen. Dabei ist ein externer und interner Informationsbedarf durch ein BUIS zu befriedigen, der je nach Funktion und Ziel des Adressaten sehr unterschiedlich sein kann. Zur Informationsbeschaffung muss das BUIS dabei auf betriebsinterne und -externe Datenquellen zurückgreifen und ggf. die Güte und Qualität dieser Daten im Rahmen der Informationsaufnahme bewerten. Ferner hat ein BUIS die Aufgabe, Funktionen und Verfahren im Rahmen des Umweltmanagements zu dokumentieren sowie weitere gesetzlich relevante Informationen bereitzustellen, wie z.B. Gesetzestexte, Genehmigungsbescheide, Umweltmanagementhandbücher, Informationen zu Gefahrstoffen, Anlagenkataser etc. [Wo05], [Hi95].

3.1 Zentrales Reporting

In den Bereich des zentralen Reporting fallen diejenigen Aktivitäten, die meist für die externe, aber auch interne betriebliche Berichterstattung notwendig sind. Dies umfasst sowohl die Sammlung von Daten aus den verschiedenen Standorten und Bereichen im Unternehmen als auch die generelle Unterstützung des betrieblichen Nachhaltigkeitsmanagements.

Für das Einsammeln der umweltrelevanten Daten werden die verschiedensten Systeme verwendet. So können sowohl softwaretechnische Eigenentwicklungen als auch eingekaufte Softwarewerkzeuge zum Einsatz kommen. In der Automobilindustrie werden beispielsweise verschiedenste Softwarewerkzeuge eingesetzt. Bei der Daimler AG wurden beispielsweise auf der Basis des intranetbasierten DUDIS (Daimler Umweltdaten Informationssystem) die umweltrelevanten Daten der weltweiten Standorte für die Berichterstattung erfasst [HW09]. Ähnliches findet sich auch bei der Volkswagen AG, die mittels ihres eigenen Umweltinformationssystems u.a. diese weltweite Datenerfassung intranetbasiert durchführt [EP09].

[2] http://www.ecointense.de/de/ecowebdesk.html

Der Bereich der Nachhaltigkeitsmanagementsysteme geht über das reine Dateneinsammeln hinaus. Hier werden aus verschiedensten Quellen Daten, auch die bereits vorgestellten Umweltdaten, eingesammelt, aggregiert und aufbereitet. Nur so lassen sich die Ansprüche der Stakeholder und Shareholder der Unternehmen effizient diesbezüglich befriedigen. Dies bedeutet u.a., auf folgenden Kommunikationswegen Daten und Informationen bereitstellen zu können:

- Nachhaltigkeits-Ratings / -Rankings
- Geschäftsbericht
- Nachhaltigkeitsbericht
- Stakeholder-Dialog
- Ad-hoc-Anfragen

Hierfür finden sich am Markt die verschiedensten kommerziellen Systeme, die ein unterschiedliches Leistungsspektrum haben. Diese sind z. B. Credit360, Enablon SD, SAP Sustainability Performance Management, SoFi Software by PE International, Tofuture CSM und WeSustain ESM [Up14].

3.2 Katastersysteme

Katastersysteme sind Informationssysteme über Objekte (z.B. Anlagen, Maschinen, Gebäude, Lager, Stoffe etc.) mit Ortsbezug. Sie sind zum Teil gesetzlich gefordert. So gibt es Kataster in den Unternehmen z.B. für Gefahrstoffe, VAwS-Anlagen, genehmigungsbedürftige Anlagen, Altlasten, Emissions- und Lärmquellen [HW09]. Sie haben zumeist folgende Aufgaben:

- Verwaltung der umweltrelevanten Anlagen eines Standortes zur Sicherstellung des ordnungsgemäßen Betriebes
- Erfassung und Dokumentation der behördlichen Genehmigungen und Auflagen
- Terminverfolgung zur Unterstützung der betrieblichen Aufgaben (z.B. Messungen und Prüfungen)
- Instrument für die jährlich wiederkehrenden Standortzertifizierungen nach EMAS (EG Öko-Audit- Verfahren)

In den Unternehmen finden sich hierfür sowohl Eigenentwicklungen als auch Standardprogramme, wie z.B. Auriga[3]. Zielstellung der Systeme ist es, folgende Punkte im Blick zu behalten, um das Umweltmanagement umfassend mit allen nötigen Informationen zu versorgen [In14]:

[3] http://www.inplus.de/produkte/auriga-professional/

- Anlagen, Teilanlagen und Betriebsmittel,
- Stoffe, Mengen, Emissionen,
- Verantwortliche Mitarbeiter, Termine und Dokumente.

3.3 Lokale Datenerfassung

Insbesondere bei größeren Unternehmen werden stoffliche Verbräuche mittels lokaler Sensoren direkt ausgelesen und in speziellen Softwarewerkzeugen weiterverarbeitet. Dies ist insbesondere stark verbreitet im Bereich der Energiedatenerfassung von Anlagen als potenzielle Erweiterung der Betriebsdatenerfassung. Es werden hier auch die verschiedensten Auswertungen und maßgeschneiderte Berichte generiert, aus denen sich zudem weitere Optimierungspotenziale ableiten lassen.

Es finden sich in diesem Bereich sowohl Standardsoftwarelösungen der großen Automatisierungstechnikanbieter als auch Eigenentwicklungen[4]. Letztendlich müssen jedoch alle Systeme die Informationen von verbauten Sensoren in den Anlagen abgreifen. Diese Dichte der Sensoren muss kontinuierlich in den Unternehmen ausgebaut werden, um zukünftig alle Anlagen im Detail abbilden zu können. In Abb. 2 wird das Beispiel eines solchen Gesamtsystems der Firma Siemens dargestellt.

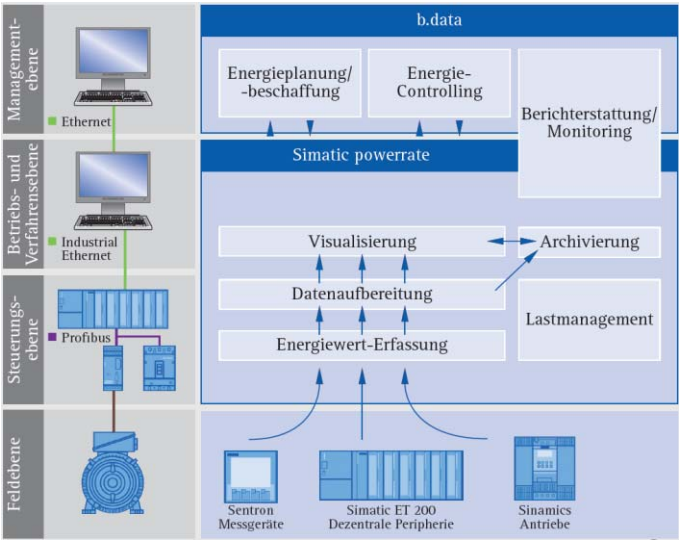


Abb. 2: Energiedatenerfassung von der Firma Siemens [SK10]

[4] Dies zeigen die Ausstellerlisten auf den einschlägigen Messen, wie z.B. die EnergieEffizienz-Messe Frankfurt oder der Teilbereich Automationstechnologien & IT auf der Hannover Messe Industrie.

Neben der Erfassung umweltrelevanter Daten durch Smart Meter spielen insbesondere für KMU auch immer häufiger mobile Erfassungsmethoden eine Rolle, um einen Medienbruch von händisch erfassten Daten zu vermeiden. Es wird dabei häufig auf moderne Smartphones und Tablets zur Erfassung umweltrelevanter Daten gesetzt, da KMU i.d.R. aus Kostengründen nicht über Smart Meter zur Erfassung von Energiedaten verfügen [Kr13]. Auch sollen damit nicht nur Energieströme, sondern allgemeiner alle potentiell wichtigen Stoffströme mobil erfasst werden können. Der Einsatz dieser modernen Technologien soll außerdem die Akzeptanz der Benutzer steigern, da sie diese bereits aus dem Privatbereich kennen und dadurch mit der Bedienung vertraut sind. Diese umweltrelevanten Daten bilden zusätzlich zu den in anderen betrieblichen Informationssystemen vorhandenen Daten eine wesentliche Informationsquelle für die erfolgreiche Durchführung von computergestützten Ressourceneffizienzmaßnahmen. Deswegen sollten BUIS über entsprechende Importschnittstellen verfügen [Wo14].

3.4 Wissensmanagement

Der Bereich des Wissensmanagements ist ebenfalls wichtig für den betrieblichen Umweltschutz. Darunter fällt die Bereitstellung von Informationen zur Unterstützung der Umweltexperten im Unternehmen als auch die Bereitstellung von Informationen und Wissen für alle Mitarbeiter. So bildet u.a. die Aufbereitung und Verfügbarmachung des betrieblichen Wissens eine wichtige Basis für einen im Unternehmen gelebten betrieblichen Umweltschutz.

Einen wesentlichen Teil im Kontext des Wissensmanagements im Umweltbereich machen jedoch die Informationen über aktuelle gesetzliche Vorschriften und Vorgaben aus. Denn der Aspekt der Rechtsicherheit hat im Umweltbereich einen hohen Stellenwert. Dies wird sichergestellt u.a. durch die Einbindung etablierter externer Angebote und Lösungen, z.B. über das Internetangebot von Umwelt-Online [HW09].

Neben diesen eher externen Angeboten gibt es auch interne Lösungen für das Wissensmanagement oder zur Mitarbeitersensibilisierung. Ein Beispiel ist der Energie- und Umweltberater der Volkswagen AG. Dies ist ein Flash-basiertes Intranet-Tool, das den Querschnitt einer Fabrik zeigt und den Nutzer intuitiv in den jeweiligen Produktionsbereichen an das Thema Umwelt und Energie heran führt und wichtige Informationen zu den Bereichen vermittelt. Gleichzeitig werden alle wichtigen Dokumente und Unterlagen für das Thema hierüber zentral bereitgestellt [Vo14a].

4 Bewertung und Entscheidungsunterstützung

Zur Vorbereitung von Entscheidungen sind die erfassten betrieblichen Umweltauswirkungen zu entscheidungsrelevanten Informationen zu verdichten. In diesem Kontext finden stoffliche, ökologische Betrachtungen immer mehr Einzug in die betrieblichen Entscheidungsprozesse. Folglich steigt hier der Bedarf, die etablierten betrieblichen

Prozesse umweltorientiert zu unterstützen. Zusätzlich muss ein BUIS in diesem Zusammenhang auch die Suche nach ökologischen Schwachstellen und die Erarbeitung von Alternativen ermöglichen [Wo05], [Hi95].

4.1 Stoff- und Energiestromanalysewerkzeuge

Unter dem Begriff Analysewerkzeuge werden hier alle Softwarewerkzeuge verstanden, die eine Ermittlung, Identifikation und Visualisierung der betrieblichen Stoff- und Energieströme erlauben. Zusätzlich können diese Ströme auch bewertet und simuliert werden, d.h. What-if-Analysen für zukünftige oder geplante Produktionssituationen durchgespielt werden. Diese Werkzeuge können bei der Ökobilanzierung und bei der Simulation von Produktionssystemen aber auch ganz allgemein für jede Analyse mit stofflichen oder energetischen Bezug eingesetzt werden. Ein Beispiel hierfür ist die Software e!Sankey[5] zur Visualisierung der Stoff- und Energieströme in Form von Sankey-Diagrammen.

4.2 Ökobilanzierung

Der Bereich der Ökobilanzierung erfordert auf Grund seiner Komplexität eigene Systeme. Anwendung findet die Ökobilanzierung in Unternehmen mit einem eher rein produktorientierten Fokus, d.h. bei der Produktentwicklung sowie der nachfolgenden Produktbewertung und zur Beurteilung der gesamten Wertschöpfungskette [BW09].

Es gibt für die ökologische Bewertung Standardsoftwarewerkzeuge, die in der Industrie Anwendung finden. Diese sind z.B. GaBi[6] oder Umberto[7]. Neben diesen Softwarewerkzeugen zur ganzheitlichen Unterstützung des Prozesses der Ökobilanzierung werden in diese Kategorie auch die verschiedenen Datenbanken und Datenquellen eingeordnet, welche die relevanten Daten zur ökologischen Bewertung der Prozesse liefern. Dies ist wichtig, um die nicht im Unternehmen stattfindenden Prozesse (sog. Mantelsysteme) mit in der ökologischen Bewertung berücksichtigen zu können.

Als Spezialfall der Ökobilanzierung ist noch das Thema Carbon Footprint zu erwähnen. Diese Auswertung als vereinfachte Ökobilanz kann sowohl von den bereits erwähnten Softwarewerkzeugen vorgenommen werden, als auch von einfacheren, darauf spezialisierten Softwarewerkzeugen.

4.3 Simulation von Produktionssystemen

Neben den klassischen Einsatzgebieten der BUIS, die sich mehr auf die Sicherstellung der Rechtskonformität, die Unterstützung des Umweltmanagements und die Bereitstel-

[5] http://www.e-sankey.com/de/
[6] http://www.gabi-software.com/
[7] http://www.umberto.de/

lung von Umweltinformationen fokussieren, gibt es einen zusätzlichen wichtigen Anwendungsbereich, nämlich die Simulation von Produktionssystemen zur Fertigungsplanungsunterstützung. Dabei geht es um die Unterstützung der bereits vorhandenen Planungsprozesse im Unternehmen durch Bereitstellung zusätzlicher umweltrelevanter Informationen. Damit sollen Umweltthemen automatisch mit in die Prozesse integriert werden. Dies kann sowohl durch komplett neue Softwarelösungen oder durch die Erweiterung der bestehenden Softwarewerkzeuge erfolgen. Nachfolgend werden einige Beispiele vorgestellt.

Beispiel: Umberto-Modell in der Lackplanung bei der Volkswagen AG

Bei der Volkswagen AG wurde ein hochkomplexes Umberto-Modell erstellt, um eine frühzeitige Planungsunterstützung bei neuen Lackieranlagen, einen Technologievergleich verschiedener Lackierprozesse und eine Simulation der Stoffströme einer Lackiererei bei veränderten Prozessparametern zu ermöglichen. Zur Realisierung dieses speziellen Anwendungsfalles wurde die Software Umberto verwendet [WB12].

Die Detaillierung des geschaffenen Modells ist in der nachfolgenden Abbildung dargestellt. So wurden in dem Modell die wesentlichen Prozesse der Lackiererei in Subnetze zerlegt und alle relevanten Stoff- und Energieströme als sog. Sankey-Diagramm visualisiert. Die wesentlichen Unterprozesse wurden in diesen Subnetzen als Transitionen erstellt und durch die Integration von Parametern flexibel gestaltet. Zur vereinfachten Bedienung wurde ein Assistent mit geführter Benutzeroberfläche zur Eingabe der wesentlichen Parameter implementiert. Dies erleichtert den Zugang zu den vielen Parametern des Modells [WB12].

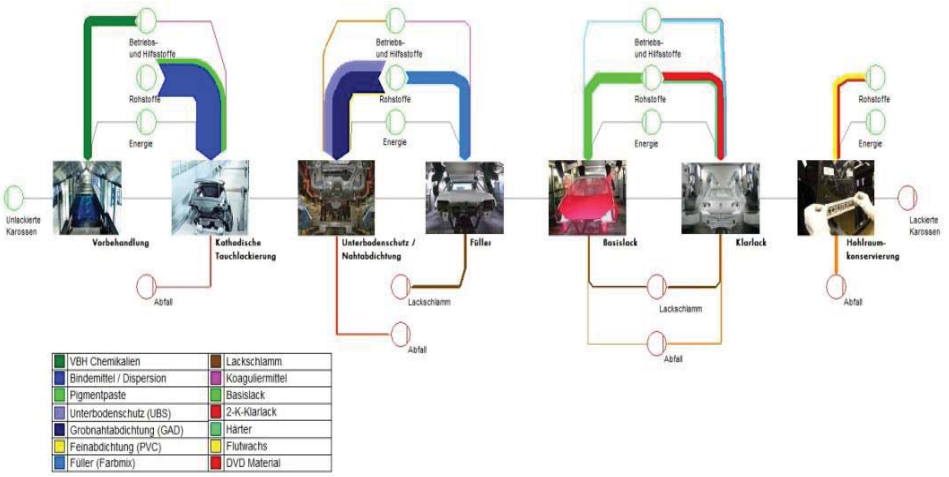


Abb. 3: Detaillierungsebenen des Umberto-Modells [WB12]

Beispiel: Plant Simulation

Als erster Vertreter und einer der wichtigsten im Kontext der betrieblichen Lösungen aus dem Bereich der Digitalen Fabrik sei hier das Softwarewerkzeug Plant Simulation[8] erwähnt. Als Programm zur Ablaufsimulation ist es ein wesentlicher Bestandteil bei der Planung neuer Fertigungsprozesse zur Absicherung der Prozesse und der Prozessfolge. Es findet hierbei zusätzlich Anwendung bei Analysen zur Prozesskettenverkürzung, Fahrweisenoptimierung und Prozessdimensionierung.

Über den Verband der Automobilindustrie (VDA) wurde ein Baukasten zur Erweiterung von Plant Simulation entwickelt und kontinuierlich erweitert. In diesem Kontext wurde auch eine in der Automobilindustrie standardisierte Methodik für die Erstellung von Simulationsmodellen mit fokussiert. Dieser VDA Automotive Bausteinkasten ist die Basis für nahezu alle Simulationsprojekte mit Plant Simulation in der Automobilindustrie [MP10].

Als neuste Entwicklung ist es nun auch möglich, Energie- und Medienverbräuche mit in die Simulation für die Zustände der Anlagen zu integrieren. Hierzu wurde der VDA Automotive Bausteinkasten um ein sog. Consumption-Modul erweitert.

5 Steuerung, Planung und Kontrolle

Das BUIS hat grundsätzlich Steuerungs-, Planungs- und Kontrollaufgaben für alle Maßnahmen des betrieblichen Umweltschutzes auf den verschiedenen Entscheidungsebenen durchzuführen. In diesem Zusammenhang gilt es u.a. die operativen Abläufe zu unterstützen, wie z.B. die Überwachung von Emissionen und die Erstellung von Entsorgungsnachweisen. Auch muss der kontinuierliche Verbesserungsprozess (KVP) im Rahmen des Umweltmanagementsystems unterstützt werden. Dies bedeutet, dass aufbauend auf einer Ist-Analyse Verbesserungsmaßnahmen generiert und umgesetzt werden müssen. Diese sind abschließend mit Hilfe der BUIS mittels eines Soll-Ist-Vergleichs einer Erfolgskontrolle zu unterziehen [Wo05], [Hi95].

5.1 Systeme zur Unterstützung des Umweltmanagementsystems

Im Kontext der Steuerung, Planung und Kontrolle gibt es Softwarewerkzeuge, die die expliziten Prozesse des Umweltmanagementsystems bzw. den dahinter liegenden Workflow unterstützen. Dazu zählen z.B. die am Markt vorhandenen Softwarelösungen EcoWebDesk oder Auriga. Sie unterstützen u.a. folgende Prozesse:

- Einheitliche Überwachung und Steuerung der Unternehmensprozesse in den Bereichen Abfall, Energie, Wasser/Abwasser etc.

[8] http://www.plm.automation.siemens.com/de_de/products/tecnomatix/plant_design/plant_simulation.shtml

- Unterstützung des kontinuierlichen Verbesserungsprozess (KVP)
- Management der internen und externen Auditierungen
- Einhaltung der gesetzlichen Vorlagen und Erfüllung von Normvorgaben
- Integriertes Dokumentenmanagementsystem
- Unterstützung der notwendigen Prozesse im Gefahrstoffmanagement und bei der Arbeitssicherheit

Auch viele der bereits unter dem Punkt Informationsbeschaffung und Dokumentation vorgestellten Systeme unterstützen das Umweltmanagement. An dieser Stelle soll jedoch die explizite Unterstützung der Prozesse im Umweltmanagementsystem bei der Steuerung, Planung und Kontrolle hervorgehoben werden.

5.2 Softwarewerkzeuge für den operativen betrieblichen Umweltschutz

Im Bereich des operativen betrieblichen Umweltschutzes finden sich die klassischen Anwendungen von BUIS. Sie sind Teil des betrieblichen Umweltmanagementsystems, können aber auch als einzelne Bereiche wahrgenommen werden. Hierzu gehören das Gefahrstoffmanagement oder auch das Abfallmanagement. Zur Unterstützung dieser vielen verschiedenen Anwendungsgebiete gibt es bereits etablierte Standardlösungen, wie z.B. EcoWebDesk und WAUplus[9].

Aber es gibt auch Softwarelösungen, die i.d.R. zur Gewährleistung der Rechtssicherheit nur einzelne Aspekte bzw. Bereiche fokussieren. Als Beispiel für den Bereich des Abfallmanagements lässt sich TRIAS AMOS[10] erwähnen. Diese Software wird u.a. bei den meisten deutschen Unternehmen der Automobilbranche eingesetzt [AT14]. Dieses Softwarewerkzeug bildet den gesamten Prozess und die gesetzlich geforderte Abwicklung und Dokumentation von der Entstehung bis zur Entsorgung der Abfälle ab. Dazu werden folgende Prozesse unterstützt:

- Bestandsführung - Lagerverwaltung - Gefahrgut-Informationen
- Entsorgungsvorgänge
- Entsorgungsdisposition
- Elektronische Nachweisführung
- Interne Kostenzuordnung
- Rechnungserfassung und Rechnungsprüfung
- Verkäufe und Ausgangsrechnungen

[9] http://www.kisters.de/wauplus.html
[10] http://www.athos.com/produkte/trias/trias-amos-fuer-erzeuger/

- Auswertung und Statistik sowie Bilanzen und Konzepte

5.3 Maßnahmenverfolgung

Ein weiterer Bereich ist die Maßnahmenverfolgung. Hierfür gibt es sowohl Eigenentwicklungen als auch erwerbbare Software. Ein Beispiel aus der Industrie für eine Eigenentwicklung ist das interne Wissensmanagement und die generelle Maßnahmenverfolgung bei der Volkswagen AG mittels der webbasierten Software Maßnahmen@web. Hierin werden alle Maßnahmen, insbesondere auch Umweltmaßnahmen, je Standort und Bereich mit einer detaillierten Beschreibung eingetragen. Diese Maßnahmen werden von der ersten Idee bis hin zur Umsetzung verfolgt. Durch dieses offene System sind die Ideen und Maßnahmen aller Standorte für alle Volkswagenmitarbeiter sichtbar [Vo14b].

6 Zusammenfassung und Entwicklungspfade

Anhand obiger Aufstellung ist zu konstatieren, dass BUIS heute hauptsächlich zur Sicherstellung der Rechtskonformität, der Unterstützung des Umweltmanagements sowie der Informationssammlung und -bereitstellung eingesetzt werden. Die BUIS in den genannten Bereichen erfüllen ihre zumeist umweltaspektbezogenen Aufgaben gut und haben sich zur Bearbeitung der genannten Problembereiche in den Betrieben etabliert. Es handelt sich in der Regel jedoch um Insellösungen und Stand-Alone-Systeme mit vergleichsweise geringer Nutzerzahl [HW09]. Auch unterstützen diese BUIS eher operative Entscheidungsfragestellungen.

Bei den betrachteten BUIS wurde explizit der Bereich der Office-Werkzeuge ausgeklammert. Lösungen auf dieser Basis haben zumeist nur einen singulären Charakter für einzelne kleine Anwendungen und keine weitere Integration in die betriebliche Infrastruktur. Sie fallen daher eigentlich nicht unter die aus der Wirtschaftsinformatik bekannte Definition von Informationssystemen [Kr11]. Nichtdestotrotz werden sie in der betrieblichen Praxis oft für kleine dezidierte Fragestellungen verwendet, die kein umfassendes Informationssystem benötigen, und daher hier im Kontext der IT-Unterstützung des betrieblichen Umweltschutzes genannt werden müssen. Aus diesem Grund schließt die oben genannte BUIS-Definition Office-Werkzeuge mit ein.

Es lassen sich aber auch zwei Entwicklungspfade ausmachen, die bisher wenig in der Fachliteratur thematisiert wurden, nämlich dass immer mehr BUIS-Funktionalitäten in anderen betrieblichen Anwendungssystemen bereitgestellt werden und teilweise taktische und strategische Entscheidungsfragestellungen unterstützen. Ein Beispiel stellen die Softwarewerkzeuge der Digitalen Fabrik dar. Diese haben bisher eher den Fokus auf eine betriebswirtschaftliche Optimierung der Produktionsprozesse und -abläufe, werden aber aktuell gerade in der Automobilindustrie in Bezug auf eine inputbezogene Abbildung von stofflichen und energetischen Verbräuchen in der Produktion erweitert, um so auch stoffliche Aspekte bei taktischen und strategischen Entscheidungsfragestel-

lungen zu berücksichtigen [Kr11]. Diese Tendenz ist zu begrüßen, da dadurch neben den klassischen Steuerungsparametern der Produktionssimulation (Durchsatz, Auslastung, Bestände) nun auch stoffliche Aspekte (Material- und Energieverbräuche) bei taktischen und strategischen Problemstellungen berücksichtigt werden können. Die dabei entstehenden Lösungen zur Unterstützung des betrieblichen Umweltschutzes in Kooperation mit den Softwarelieferanten oder in Forschungsprojekten können letztendlich förderlich für die gesamte produzierende Industrie sein, wenn es gelingt, diese einzelnen Dateninseln zu vernetzen und mit der bestehenden betrieblichen IKT-Infrastruktur zu integrieren.

Bei näherer Betrachtung der Unternehmen und ihrer derzeitigen BUIS-Landschaft zeigt sich, dass insbesondere im Bereich der umweltorientierten Produktion folgende Punkte wesentlich sind für die zukünftige Entwicklung von BUIS:

- Ganzheitlich Ausbau der Datenerfassung aller wesentlichen Energie- und Stoffströme im Betrieb,
- Integration von Energie- und Umweltaspekten in die bestehenden Softwarewerkzeuge und Prozesse im Unternehmen,
- Nutzung der Möglichkeiten der Digitalen Fabrik als wesentlicher Hebel zur Integration von Nachhaltigkeitsaspekten bei taktischen und strategischen Planungsfragestellungen.

Es lässt sich somit festhalten, dass zumindest in Großunternehmen doch eher auf eine Anpassung, Erweiterung und Vernetzung bestehender IKT-Lösungen zum Vorteil des betrieblichen Umweltschutzes gesetzt wird, als dass die Neuentwicklung von IKT-Lösungen mit kompletter integrierter Unterstützung sämtlicher betrieblicher Umweltschutzaspekte vorangetrieben wird. Allerdings werden wohl die existieren BUIS zur Unterstützung einzelner dedizierter Aspekte des betrieblichen Umweltschutzes, wie z.B. dem Stoffstrommanagement oder Abfallmanagement, weiterhin Bestand haben.

Literaturverzeichnis

[AT14] ATHOS GmbH: Referenzen des ATHOS Geschäftsbereichs TRIAS http://www.athos.com/unternehmen/referenzen/ (Letzter Zugriff: 20.08.2014).

[Be02] Beucker, S. et. al..: Betriebliche Umweltinformationssysteme und ihre Funktion für die Ressourceneffizienzrechnung. Ergebnisbericht 2.3 des Forschungsberichtes CARE, Stuttgart, 2002.

[Bo13] Boehnke, B., Möller, A., Wohlgemuth, V.: Integration of Material Flow Management into Company Processes within the Automotive Industry. In: Page, B., Fleischer, A., Göbel, J., Wohlgemuth, V. (Hrsg.): Environmental Informatics and Renewable Energies: 27th International Conference on Informatics for Environmental Protection.. Shaker, Aachen, 2013, S. 434-442.

[BW09] Boehnke, B., Wohlgemuth, V.: Konzeption zur Integration eines ganzheitlichen Stoffstrommanagements bei der Volkswagen AG. In: Fischer-Stabel, P., Kremers, H., Susini, A., Wohlgemuth, V. (Hrsg.): Environmental Informatics and Industrial Environmental Protection – Concepts, Methods and Tools – Volume 3, Shaker, Aachen, 2009, S. 71-82.

[EP09] Ernesti, L., Petri, T.: Webbasiertes Betriebliches Umweltinformationssystem (BUIS) bei einem großen Automobilhersteller. In: Fischer-Stabel, P., Kremers, H., Susini, A., Wohlgemuth, V. (Hrsg.): Environmental Informatics and Industrial Environmental Protection – Concepts, Methods and Tools – Volume 3, Shaker, Aachen, 2009, S. 67-70.

[Hi95] Hilty, L.M.: Betriebliche und überbetriebliche Umweltinformationssysteme als informationstechnische Infrastruktur für das Stoffstrommanagement. In: Schmidt, M, Schrob, A. (Hrsg.): Stoffstromanalysen in Ökobilanzen und Öko-Audits.. Springer, Berlin u.a., 1995, S. 193-205.

[HR95] Hilty, L.M., Rautenstrauch, C.: Betriebliche Umweltinformatik. In: Page, B., Hilty, L.M. (Hrsg.): Umweltinformatik - Informatikmethoden für den Umweltschutz und Umweltforschung. 2. Auflage, Oldenbourg, München, 1995, S. 295-312.

[HW09] Heldt, K.; Wohlgemuth, V.: Typische Entwicklungs- und Entscheidungsprozesse zu betrieblichen Umweltinformationssystemen am Beispiel der Daimler AG. In: Fischer-Stabel, P., Kremers, H., Susini, A., Wohlgemuth, V. (Hrsg.): Environmental Informatics and Industrial Environmental Protection – Concepts, Methods and Tools – Volume 3, Shaker, Aachen, 2009, S. 83-90.

[In14] Inplus: Umwelt und Sicherheit - Branchenlösung für die Papierindustrie. Germering bei München, 2014.

[JL11] Junker, H.; Lang, C. V.: Problems of practical CEMIS (Corporate Environment Information Management Systems) usage. In: Golinska, P.; Fertsch, M.; Marx-Gómez, J. (Hrsg.): Information Technologies in Environmental Engineering, Environmental Science and Engineering. Springer Verlag, Berlin u.a.., 2011, S. 153-163.

[Ju10] Junker, H.: Die Beliebigkeit betrieblicher Umweltinformationssysteme – eine Polemik. In: Greve, K., Cremers, A.B. (Hrsg.): Integration of Environmental Information in Europe, 24th International Conference on Informatics for Environmental Protection, Shaker, Aachen, 2010, S. 232 – 247.

[Kl14] Klingebiel, K., Hackstein, L., Cirullies, J., Parlings, M., Hesse, K., Hohaus, C., Jung, E.-N.: Ressourceneffiziente Logistik. In: Neugebauer, R. (Hrsg.): Handbuch Ressourcenorientierte Produktion.. Carl Hanser Verlag, München, 2014, S. 719-747.

[Kr11] Krcmar, H.: Informationsmanagement, 5. Auflage, Springer, Heidelberg [u.a.], 2011, S. 27ff.

[Kr13] Krehahn, P., Ziep, T., Schiemann, L., Wohlgemuth, V.: OpenResKit – Herausforderungen und aktuelle Entwicklungstendenzen bei der software-technischen Unterstützung von Ressourcen- und Energieeffizienzfragestellungen auf der Basis einer Client-/Server-Architektur. In: Page, B., Fleischer, A., Göbel, J., Wohlgemuth, V. (Hrsg.): Environmental Informatics and Renewable Energies: 27th International Conference on Informatics for Environmental Protection, Shaker, Aachen, 2013, S.

434-442.

[La03] Lang, C., Keil, M., Heubach, D.: Stoffstrommanagement für klein- und mittelständische Unternehmen der Mineralwasserbranche – Arbeitsbericht aus dem Forschungsprojekt INTUS. Stuttgart, 2003.

[MP10] Mayer, G., Pöge, C.: Auf dem Weg zum Standard – Von der Idee zur Umsetzung des VDA Automotive Bausteinkastens. In: Zülch, G., Stock, P. (Hrsg.): Integrationsaspekte der Simulation: Technik, Organisation und Personal. S. 29-36. KIT Scientific Publishing, Karlsruhe, 2010.

[Pa89] Page, B.: Studie über DV-Anwendungen in den Umweltbereichen des Bundes und der Länder – Phase I und II. UBA-Texte 35/86 und 30/89, Springer, Marburg, 1989.

[Ra99] Rautenstrauch, C.: Betriebliche Umweltinformationssysteme – Grundlagen, Konzepte und Systeme. Springer: Berlin u.a., 1999.

[SK10] Schmidt, C., Kunzmann, G.: Intelligentes Energiemanagement. etz 3/2010, 2010, S. 2-4.

[Up14] Uphoff, K., Solsbach, A., Van Vliet, S., Rapp, B., Isenmann, R.: Nachhaltigkeitsberichterstattung – Anforderungen zum Datenaustausch aus Sicht der Praxis. In: Wohlgemuth, V., Lang, C.V., Marx-Gomez, J. (Hrsg.): Konzepte, Anwendungen und Entwicklungstendenzen von betrieblichen Umweltinformationssystemen (BUIS), Shaker, Aachen , 2014, S. 205-216.

[Vo14a] Volkswagen AG: Volkswagen führt virtuellen Energieberater ein, http://www.volkswagenag.com/content/vwcorp/info_center/de/news/2011/06/energy.html (Letzter Zugriff: 20.08.2014).

[Vo14b] Volkswagen: Autogramm 8/2010 - Das Forum „Optimierung Indirekte Bereiche Produktion" arbeitet markenübergreifend / Top-Beispiele aus den Standorten http://autogramm.volkswagen.de/08_10/volkswagenweg/volkswagenweg_01.html, (Letzter Zugriff: 20.08.2014).

[WB12] Witte, S., Boehnke, B.: Parametrisiertes benutzergeführtes Simulationsmodell einer idealisierten Automobillackiererei. In: Wohlgemuth, V. (Hrsg.): Einsatz der Software UMBERTO in angewandter Forschung und Praxis – Anwendungsfälle und Praxisbeispiele des UMBERTO Competence Center Berlin (UCC Berlin), Shaker, Aachen, 2012, S. 99-120.

[Wo05] Wohlgemuth, V.: Komponentenbasierte Unterstützung von Methoden der Modellbildung und Simulation im Einsatzkontext des betrieblichen Umweltschutzes – Konzeption und prototypische Entwicklung eines Stoffstromsimulators zur Integration einer stoffstromorientierten Perspektive in die auftragsbezogene Simulationssicht. Dissertation, Universität Hamburg, Hamburg, 2005.

[Wo14] Wohlgemuth, V., Krehahn, P., Ziep, T., Schiemann, L.: Entwicklung eines Open-Source basierten Baukastens zur Unterstützung und Etablierung der Ressourceneffizienz in produzierenden KMU. In: Wohlgemuth, V., Lang, C.V., Gomez, J.M. (Hrsg.): Konzepte, Anwendungen und Entwicklungstendenzen von betrieblichen Umweltinformationssystemen (BUIS), Shaker, Aachen, 2014, S. 41-58.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Integration und Weiterentwicklung bestehender Energiemanagement-Applikationen mit dem OpenResKit-Framework

Johannes Boß[1] und Volker Wohlgemuth[2]

Abstract: Für kleine und mittlere Unternehmen gerät das Thema ressourceneffiziente Produktion vermehrt in den Fokus. Der Einsatz von Informationssystemen ist hier unabdingbar, und viele Unternehmen können bereits auf eine Datengrundlage, jedoch oftmals in unterschiedlicher Qualität, zurückgreifen. Standardprodukte können diese Daten nicht integrieren und sind für KMU zu teuer. Dieser Beitrag beschreibt den Versuch, bestehende Informationsressourcen und Anwendungen im Unternehmen zu nutzen. Am Beispiel des Energiemanagements wird einerseits eine zentrale Energiemanagementanwendung weiterentwickelt und andererseits im Unternehmen bereits vorhandene Daten hierin integriert. Die Umsetzung erfolgte mit dem an der HTW Berlin entwikkelten OpenResKit-Framework. Besonders vorteilhaft erweist sich die modulare Architektur mit einer leicht anpassbaren Datenbank und die nicht proprietären Austauschformate, um die Daten clientunabhängig verfügbar zu machen.

Keywords: BUIS, Energiemanagement, ISO 50001, OpenResKit-Framework

1 Einleitung

Kleine und mittlere Unternehmen (KMU) können mit einer ressourceneffizienten Produktion viele brachliegende Potentiale nutzen [Vd11, S. 4]. Die Motivation für ressourceneffiziente Produktion kann von unterschiedlicher Art sein. Das vordergründige Argument ist die Unterstützung des betrieblichen Umweltschutzes und die damit verbundene Verringerung der Emission von Treibhausgasen. Wirtschaftliche Vorteile sind aus unternehmerischer Sicht jedoch häufig gewichtigere Gründe. Steigende Energiekosten können durch einen sparsameren Energieverbrauch abgefedert werden. Durch die planmäßige Suche nach Verbesserungspotentialen im eigenen Energieverbrauch und die Investition in energieeffiziente Technologien sinken die Energieverbräuche im Unternehmen. Auch der Staat unterstützt Unternehmen mit verschiedenen Fördermaßnahmen. Hier ist die Befreiung von der Umlage nach dem Erneuerbare-Energien-Gesetz (EEG-Umlage) für energieintensive Unternehmen zu nennen (§ 64 Abs. 1 EEG 2014), oder die Steuervergünstigung zum Spitzenausgleich im Rahmen der Energie- und Stromsteuer, zu deren Voraussetzungen ein Energiemanagement im Unternehmen gehört (§ 10 Abs. 3 StromStG). Die Anstrengungen zur Energieeffizienz können nach außen wirksam durch eine Zertifizierung dargestellt werden. Dies kann für die öffentliche Wirkung generell

[1] HTW Berlin, Fachbereich 2, Wilhelminenhofstraße 75A, 12459 Berlin, joha.boss@gmail.com
[2] HTW Berlin, Fachbereich 2, Wilhelminenhofstraße 75A, 12459 Berlin, volker.wohlgemuth@htw-berlin.de

sinnvoll sein, oder durch einen Auftraggeber explizit gefordert werden. Die strukturierte Analyse und Verbesserung der eigenen Unternehmensprozesse, die häufig von externen Fachleuten begleitet wird, ist ein Nebeneffekt der Anstrengungen.

Ein systematischer Ansatz für eine energieeffizientere Produktion ist die Einführung eines Energiemanagementsystems, beispielsweise nach DIN ISO 50001. Die praktische Umsetzung der Dokumentations- und Überwachungspflichten erfolgt mit einem Softwaresystem. Der nächstliegende Ansatz ist die Einführung eines Standardsoftwareproduktes zur Unterstützung des Energiemanagements. In der Praxis der KMU stehen dem einige Vorbehalte entgegen, falls die hohen Anschaffungskosten überhaupt einen Kauf zulassen. Die Funktionalität vieler Produkte ist vom Umfang nicht auf KMU angepasst und überfordert die Fachabteilungen, die geringere Anforderungen an das System haben. Ein zusätzliches Produkt vergrößert die Komplexität der bestehenden IT-Anwendungslandschaft im Unternehmen [Ti13, S. 19]. Bei Standardprodukten müssen die Schnittstellen zu den besonderen Gegebenheiten in einem einzelnen Unternehmen, falls überhaupt möglich, erst konfiguriert werden [Ba09, S. 14]. Dieser Beitrag verfolgt deshalb einen Bottom-Up Ansatz, welcher versucht, einerseits vorhandene Daten im Unternehmen zu nutzen und andererseits eine bereits eingesetzte Anwendung weiterzuentwickeln. Innerhalb der Software sollen alle, das Energiemanagement betreffende Aufgaben verwaltet werden können. Geschäftsprozesse, die bisher mit Excel-Dokumenten oder veralteten Datenbanken durchgeführt wurden, werden in der neuen Anwendung abgebildet und alte Datenbestände, wenn möglich migriert. Diese pragmatische Vorgehensweise wird bei einem Partnerunternehmen validiert.

2 Grundlagen

In der Arbeit werden Elemente zweier Ansätze der Softwaretechnik zur Evolution von Softwaresystemen verwendet. Die erfolgversprechendste Option zur Abdeckung zusätzlicher IT-Bedarfe ist laut Huber, „bestehende Anwendungen durch Erweiterungen und Änderungen an neue Anforderungen anzupassen.“ [Hu14, S. 33] Dies gilt insbesondere dann, wenn es sich bei der Anwendung um eine Eigenentwicklung handelt und im Unternehmen noch ausreichend Wissen verfügbar ist. Dieser Ansatz wird auch als Softwareevolution bezeichnet [Sn12]. Die Prozesse der Informations- und Datenintegration helfen, Komplexität der Anwendungslandschaft zu reduzieren und ermöglichen es, auf die bestehenden Daten zuzugreifen.

2.1 Softwareevolution

Die Begriffe Softwareevolution, Wartung, Pflege oder Veränderungsmanagement werden uneinheitlich verwendet und sind abhängig vom verwendeten Vorgehensmodell und der Betrachtungsweise. Hier wird auf das Staged Model of the Software Lifecycle von Bennett et al. [BRW02] zurückgegriffen, das in Abbildung 1 dargestellt ist. Hier gibt es nach einer initialen Entwicklungsphase eine Evolutionsphase, in der die Software um

einen größeren Funktionsumfang erweitert wird und Anforderungen korrigiert werden. Danach folgt eine Servicephase zur Fehlerbehebung und zur Anpassung minimaler Funktionalitäten, bevor in der Abwicklungs- beziehungsweise der Beendigungsphase das Produkt stillgelegt wird.

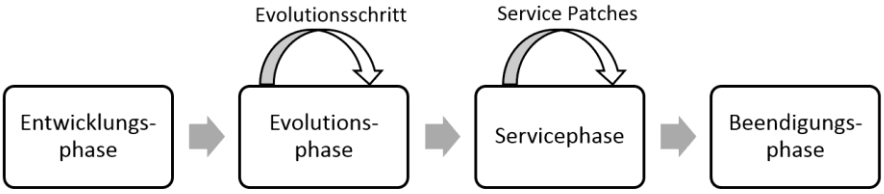


Abb. 1: Staged Model of the Software Lifecycle (vgl. Sommerville [So11, S. 236])

In [BF14] werden einige Techniken, die bei Evolutionsprojekten angewendet werden, beschrieben und nachfolgend umrissen.

- Softwareanalyse: Zunächst muss der Versuch unternommen werden, den Quellcode zu lesen und zu verstehen. Eine Dokumentation der Architektur kann hilfreich sein, ist aber oft nicht vorhanden. Es gibt einige Tools, die die Codeanalyse unterstützen. Viele Entwicklungsumgebungen verfügen mittlerweile über solche Werkzeuge.
- Reengineering: Unter Reengineering versteht man die Überprüfung und Verbesserung kompletter Komponenten. Die Funktionalität bleibt vollständig erhalten. Ziel ist, mittels einer verbesserten Organisation des Codes die Wartbarkeit zu verbessern. Davon abzugrenzen ist das Refactoring, das zwar aus gleichen Gründen durchgeführt wird, aber nur kleinere Teileinheiten des Quellcodes (Methoden, Variablen) umfasst und mit modernen Entwicklungswerkzeugen (teil-)automatisiert durchgeführt werden kann.
- Reverse Engineering: Als Reverse Engineering wird die Analyse eines Programmes mit dem Ziel einer vollständigen Neuentwickelung ganzer Komponenten bezeichnet. Gründe können die Verbesserung der Softwarequalität, die Anpassung an neue Plattformen, oder das Ersetzen von Altsystemen (Legacy Code) sein. Auch hier bleibt die Funktionalität gleich. Von steigender Bedeutung ist das Reverse Engineering von Datenbanken. Hierbei wird auf Basis einer physischen Datenbank das Datenbankschema ermittelt. Eng mit dem Reverse Engineering verbunden ist auch die Nachdokumentationen.
- Migration: Unter Migration wird die Portierung einer Software, oder Teile davon, auf andere Plattformen verstanden. Das kann zum Beispiel die Änderung einer unterliegenden Datenbank oder der Ersatz eines nicht mehr unterstützen Frameworks sein.

- Retirement: Am Ende des Softwarelebenszyklus steht das Abschalten der Software. Die Auswirkungen sind zu überprüfen, wie zum Beispiel die Zugänglichkeit von Archivdaten, und der Ersatz ist sicherzustellen. Das Auslaufen einer Software ist eng mit der Migration verknüpft.

2.2 Informations- und Datenintegration

Im Unternehmensalltag gibt es häufig eine fragmentierte, heterogene Anwendungslandschaft [Ti13, S. 71]. Die Daten liegen in unterschiedlichen Formaten, Medien und Strukturen vor und stehen für einen unternehmensweiten Gebrauch nicht zur Verfügung [Ti13, S. 136]. Konsolidierungsprojekte in den Bereichen Applikationen und Daten können helfen, die Heterogenität der Anwendungen aufzulösen, aber auch um vorhandene Daten unternehmensweit nutzbar zu machen. Das IT-Management wird so vereinfacht, die Qualität der Anwendungslandschaft erhöht und Kosten reduziert[Ti13, S. 129]. Mit der Überführung von einzelnen Datensilos zu einer möglichst unternehmensweiten Informationssicht, lassen sich Mehrfacherfassungen und Datenredundanzen vermeiden. Die Daten- und Informationsintegration lässt sich auf mehreren Ebenen unterscheiden. Jung gibt eine Übersicht über verschiedenen Definitionen [Ju06]. Die Integration lässt sich einerseits durch Verbinden oder durch Vereinigen der Datenelemente erreichen. Beim Verbinden bleiben die möglicherweise redundanten Ausgangsbestände erhalten, aber es werden logische Beziehungen zwischen den Beständen auf Anwendungsebene hergestellt. Manchmal wird diese Art auch als Interoperabilität bezeichnet [PB06]. Bei der Integration durch Vereinigen werden diese Datenbestände verschmolzen und Redundanzen eliminiert.

Weitere Begriffe zielen auf die Abstraktionsebene ab. Ein Data Warehouse wird von verschiedenen unternehmensinternen und -externen Datenquellen gespeist und bereitet die Daten auf. Geschieht die Koppelung der einzelnen Informationsquellen nicht auf Datenebene, sondern auf Anwendungsebene, kann man eher von Enterprise Application Integration (EAI) sprechen [Me12, S. 71]. Hierbei wird jedes eingebundene System mit einem Adapter an einen zentralen Bus angeschlossen. Mit dem Begriff Service-oriented Architecture wird ein ähnliches Konzept beschrieben, dass sich an Geschäftsprozessen orientiert [Me12, S. 71].

Ebene	Konzept	Techniken
Daten	Data Warehouse	Datenintegration
Anwendungen	Enterprise Application Integration	Anwendungsintegration
Prozesse	Service-oriented Architecture	Geschäftsprozessintegration

Tab. 1: Integrationsformen nach Abstraktionsebene

3 OpenResKit-Framework

Die technische Umsetzung erfolgt mit dem OpenResKit-Framework. Dies ist ein Forschungsprojekt, dass an der HTW Berlin im Studiengang Betriebliche Umweltinformatik durchgeführt worden ist. Hierbei wurden Open-Source-basierende Softwarewerkzeuge zur innerbetrieblichen Ressourcen- und Effizienzfragestellungen erstellt [Wo14]. Der Nutzerfokus liegt auf kleinen und mittleren Unternehmen, die mit einfachen und problemspezifischen Applikationen die Ressourceneffizienz in ihrem Betrieb erhöhen können. Ziel war, einen zentralen Datenpool mit einer standardisierten Schnittstelle zu schaffen. Dieser wurde in Form des OpenResKit-Hubs, hier nachfolgend auch ORK-Server genannt, realisiert. Die Plattform ist leicht erweiterbar und modular verwendbar. Innerhalb des Forschungsprojektes und weiteren Projekten sind bereits einige Softwarebausteine entstanden. Darunter ist auch eine Anwendung zum Maßnahmenmanagement, die erweitert und an die Bedürfnisse des Partnerunternehmens angepasst wurde. Bei der OpenResKit-Architektur handelt sich um eine typische zweischichtige Client-Server Architektur. Innerhalb des Forschungsprojektes sind hauptsächlich Desktop-Clients und mobile Anwendungen entstanden. Für die Desktopanwendung sind einige wiederverwendbare Komponenten vorhanden, auf denen die zu erstellende Software aufbaut. Die Integrationsfragestellung dieser Arbeit rückt allerdings den Server in den Mittelpunkt.

Der Server erfüllt zwei zentrale Aufgaben. Erstens stellt er die Domänenmodelle bereit und persistiert sie in der eigenen Datenbank, üblicherweise eine Microsoft SQL Server Datenbank (Express Edition). Dabei wird das Microsoft Entity Framework (EF) als objektrelationaler Mapper (ORM) verwendet. Dieser wandelt die Klassenstrukturen in das Datenbankmodell um und bietet eine Abstraktionsschicht, die zwischen Datenbank und Anwendung vermittelt. Zweitens wird eine Webschnittstelle zum Datenaustausch, nach dem OData-Protokoll[3], angeboten, um die Clients anzubinden. Besonders erwähnenswert ist das dynamisch anpassbare Domänenmodell, dass für die große Flexibilität und Modularität verantwortlich ist. So können zur Laufzeit neue Module hinzugefügt werden. Die Datenbank reagiert mit einer entsprechenden Schemaänderung und einer automatischen Datenbankmigration. Bei der Umsetzung wurde dabei auf das Managed Extensibility Framework (MEF) zurückgegriffen. Diese Bibliothek ist Teil des Microsoft .NET-Frameworks und ermöglicht die Entwicklung erweiterbarer Anwendungen, ohne eine Konfiguration vornehmen zu müssen oder fest codierte Abhängigkeiten zu berücksichtigen. Somit lassen sich unbekannte, kompilierte Erweiterungen einbinden. Der Anwender hat damit eine einfache Möglichkeit das Datenmodell anzupassen und zu erweitern, was den Prozess der Datenintegration beschleunigt.

[3] http://www.odata.org/

4 Konzept

Basis der Softwarelösung ist der bestehende OpenResKit-basierte Client zum Maßnahmenmanagement und das dazugehörige Servermodul. Innerhalb des Partnerunternehmens, einem KMU aus der Kunststoffverarbeitung, werden im Bereich Umwelt- und Energiemanagement und der damit eng verknüpften Betriebstechnik im Wesentlichen die nachfolgend genannten Anwendungen eingesetzt. Die Anlagen und damit Energieverbraucher werden in einer MySQL-Datenbank mit verschiedenen Zugriffsclients verwaltet. Ein Raumbuch wird in einer lokalen Microsoft Access Datei geführt. Zusätzlich werden einige Auswertungen mithilfe von e!Sankey und einer damit verknüpften Microsoft Excel Tabelle gemacht. Die Anforderungen des Partnerunternehmens sehen vor, sämtliche Aufgaben im Rahmen des Energiemanagements nach DIN ISO 50001 innerhalb einer Anwendung zu bearbeiten. Deshalb sollen die vorher genannten Anwendungen in abgelöst und die Funktionalität in die OpenResKit-Anwendung Maßnahmenmanagement übertragen werden. Dies betrifft insbesondere die Informationen der Energieverbraucher und eine Raumdatenbank aus der Betriebstechnik. Zusätzlich sind zu den Energieverbrauchern Messwerte und Ablesungen zu speichern. Die zusätzlichen Informationen sollen auch für eine zeitbezogene Energieverbrauchsauswertung innerhalb der Anwendung genutzt werden, die die geschätzten Energieeinsparungen mit den tatsächlichen Betriebswerten vergleicht. Die lokale MySQL-Datenbank wird für diesen Bereich nicht mehr weitergeführt. Allerdings muss die Möglichkeit bestehen, weiterhin auch außerhalb des Clients auf die Datensätze zuzugreifen, und auch dort Veränderungen vorzunehmen. Die Daten sollen auch für weitere Anwendungen, zum Beispiel für die Sankey-Auswertungen, zur Verfügung stehen.

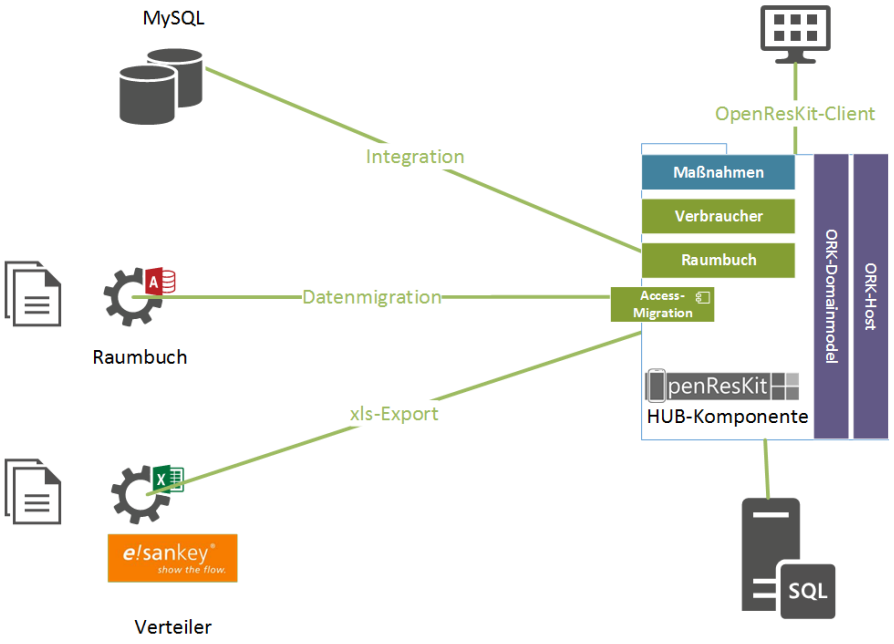


Abb. 2: Konzept für die Integration und Erweiterung der bestehenden Anwendung

5 Implementierung

Als Erstes war die Software der bestehenden Anwendung zu analysieren. Obwohl diese nach Benutzerangaben fast fehlerfrei läuft, wurde ein Reengineering durchgeführt, um die Weiterentwicklung zu ermöglichen und die Software zu verbessern. Dies beinhaltete auch das Entfernen sogenannter „Code Smells“, also schlecht strukturierter Programmcode, der die Verständlichkeit und damit die Wartbarkeit erschwert [Fo00]. Davon war besonders der Client betroffen. Für diesen entstanden in der Folge neue Module, die die zusätzlichen Kundenanforderungen abdecken. Besonderes Augenmerk sollte jedoch auf die technischen Besonderheiten gelegt werden, die mit der Erweiterung und Integration zusammenhängen.

5.1 Datenbankmigrationen

Für das Erstellen oder Ändern des Domänenmodells bietet EF zwei Ansätze an. Beim Model-First-Ansatz wird zunächst mit einem Designer ein grafisches Datenmodell erstellt. Daneben gibt es den Code-First-Ansatz, der bei OpenResKit verfolgt wurde. Es werden Klassen erstellt, die das Domänenmodell repräsentieren. Mit einer Konfigurationsdatei können Abhängigkeiten und Schlüsselbeziehungen modelliert werden. Seit

Version 4.3 bietet EF zwei Migrationsstrategien an, um in der Datenbank die Modelländerungen widerzuspiegeln. OpenResKit setzt auf die automatische Migration. Hierbei wird das bestehende Datenmodell mit der Datenbank abgeglichen und gegebenenfalls ein Update durchgeführt. Leider erkennt der automatisierte Updatemechanismus keine Umbenennungen von Tabellen oder Spalten, so dass alte Spalten gelöscht und neue erstellt werden.

Alternativ gibt es die eine codebasierte Migration. Dies widerspricht aber dem modularen Aufbau der OpenResKit-Architektur. Erst zur Laufzeit kennt der OpenResKit-Server die eingebundenen Module. Folglich können sich eventuelle codebasierte Migrationsdateien auch nicht auf die letzte Datenbankversion beziehen, da die Module das aktuelle Schema nicht kennen. Ein Update der Datenbank wäre nicht möglich. Ob sich dieses Problem mit automatisierten Merge-Migrations-Befehlen, analog zur Vorgehensweise in Teamumgebungen, lösen lässt, bedarf einer tiefergehenden Betrachtung.

5.2 Integration mittels ETL-Prozess

Eine typische Vorgehensweise zur Datenkonsolidierung ist der ETL-Prozess. ETL steht für Extract-Transform-Load, also Extrahieren-Transformieren-Laden. Zunächst werden die relevanten Daten aus den Quellsystemen entnommen. Der Prozess muss nicht einmalig ablaufen, sondern kann ereignisgesteuert, periodisch oder permanent sein. In der Transformationsphase werden die Datensätze so geändert, dass sie einheitlich in das Zielsystem passen. Dieser Teilschritt ist die aufwändigste Aufgabe, da die strukturelle, syntaktische und semantische Heterogenität der Quellsysteme überwunden werden muss. Anschließend werden die Daten in das Zielsystem geladen [Ro13, S 37 - 40]. Dieser Prozess wurde mithilfe von SQL Server Integration Services durchgeführt. Die Konfiguration mit der Entwickleroberfläche ist denkbar einfach. Es können verschiedene Datenflussaufgaben definiert werden. Sogenannte Quellen- und Zielassistenten werden angeboten, um die Verbindung zu den Ausgangsdaten beziehungsweise der Zieldatenbank herzustellen.

In dem hier beschriebenen Projekt wurden einmalig eine Access-Datei und einige Datensätze aus einer MySQL-Datenbank in die SQL Express Zieldatenbank migriert. Die Zielverbindung zur SQL-Datenbank kann mit dem Assistenten hergestellt werden. Das gewünschte Datenbankschema muss bereits vorhanden sein und wird nach dem Code-First-Ansatz erstellt. Die Quellverbindung kann mit einer OLE DB oder ODBC Schnittstelle realisiert werden. Beides sind von Microsoft entwickelte Schnittstellen, die eine einheitliche Datenbankzugriffsschicht bilden. Gegebenenfalls muss für das verwendete Quellsystem ein passender Treiber oder Connector dem Entwicklungssystem hinzugefügt werden. In diesem Projekt wurde eine MySQL Datenbank in der Version 4.1 verwendet. Ein standardmäßig nicht mehr verfügbarer Treiber konnte über die Website heruntergeladen werden.

5.3 OData Schnittstelle

Die Daten werden vom OpenResKit-Server mit einer OData-Schnittstelle angeboten. OData ermöglicht die Erstellung und Verarbeitung von REST-basierten Datendiensten. Ressourcen werden dabei über eine HTTP-Nachricht angefragt und bearbeitet. REST steht für Representational State Transfer und beschreibt Prinzipien zur Entwicklung verteilter Systeme. Antworten werden im JSON- oder ATOM-Format zurückgegeben. Da OData auf diesen weitverbreiteten Internettechnologien aufbaut, ist OData mit allen gängigen Programmiersprachen ansprechbar und es können plattformunabhängig Clients entwickelt werden. Die Datenbank kann natürlich auch weiterhin direkt mittels SQL selbst angesprochen werden.

6 Fazit

Durch das Projekt, das konsequent zusammen mit den Nutzern durchgeführt wurde, konnte die bisherige Komplexität der IT-Anwendungslandschaft verringert werden. Arbeitsschritte, die bisher in vier verschiedenen Anwendungen ausgeführt wurden, lassen sich nun in einer Anwendung bearbeiten. Eine Datenbank mit einer mehr als zehn Jahre alten Softwareversion wurde abgelöst. Die verschiedenen Datenstrukturen wurden in einer Datenbank unter Nutzung der beschriebenen Hub-Komponente vereinheitlich. Diese wird anders als die Ausgangsdatenbestände, auf einem zentralen Server redundant gesichert ausgeführt. Damit wurden aus Sicht des IT-Managements erhebliche Qualitätsverbesserungen erreicht. Die größten Vorteile ergaben sich aber aus Sicht der Nutzer. Alle relevanten Daten des Energiemanagements sind nun in einer Anwendung einsehbar und bearbeitbar. Der bisherige manuelle Datenabgleich zwischen verschiedenen Anwendungen entfällt. Weiterhin konnten aufgrund der zusätzlichen Möglichkeiten weitere Anforderungen, wie eine Auswertefunktion, umgesetzt werden. Die Daten sind über die OData-Schnittstelle auch in anderen Anwendungen, wie einer Excel-Tabelle für die e!Sankey-Auswertung, verfügbar.

OpenResKit hat sich für die Umsetzung als geeignet erwiesen. Insbesondere der modulare Aufbau des Hub-Servers bietet sich für pragmatische Umsetzung der Erweiterungen und Integrationsaufgaben an. Ein Update des Entity Frameworks auf Version 6.1 ist zu empfehlen, da hier umfangreiche Möglichkeiten zum Reengineering angeboten werden. So kann aus bestehenden Datenbanken der Quellcode generiert werden, um den Code-First Ansatz zu verfolgen. Für Integrations- und Erweiterungsaufgaben erleichtert dies die Erstellung des Domänenmodells erheblich.

Literaturverzeichnis

[Ba09] Balzert, H.: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements Engineering. Spektrum Akademischer Verlag, Heidelberg, 2009.

[BF14] Bourque, Pierre; Fairley, R. E.: Swebok: Guide to the software engineering body of knowledge. Version 3.0. Los Alamitos, CA, IEEE Computer Society, 2014.

[BRW02] Bennett, Keith H.; Rajlich, Vaclav; Wilde, Norman: Software evolution and the staged model of the software lifecycle. In: Adcvances in computers (2002), Nr, 56, S. 1-54.

[Fo00] Fowler, M.: Refactoring. Wie Sie das Design vorhandener Software verbessern. Addison-Wesley, München u.a., 2000.

[Hu14] Huber, Sebastian: Informationsintegration in dynamischen Unternehmensnetzwerken: Architektur, Methode und Anwendung. Wiesbaden, Gabler, 2014.

[Ju06] Jung, R.: Architekturen zur Datenintegration. Gestaltungsempfehlungen auf der Basis fachkonzeptueller Anforderungen, Deutscher Universitäts-Verlag/GWV-Fachverlage GmbH, Wiesbaden, 2006.

[Me12] Mertens, P. et al.: Grundzüge der Wirtschaftsinformatik. Springer Berlin Heidelberg; Imprint: Springer, Berlin, Heidelberg, 2012.

[PB06] Pellegrini, Tassilo ; Blumauer, Andreas ; Semantic Web: Wege zur vernetzten Wissensgesellschaft, 1. Auflage, Berlin, Springer, 2006.

[Ro13] Rossak, I. (Hrsg.): Datenintegration. Integrationsansätze, Beispielszenarien, Problemlösungen, Talend Open Studio; mit 15 Tabellen sowie Kontrollfragen und Aufgaben. Hanser, München, 2013.

[Sn12] Sneed, H. M.: Nachhaltigkeit durch gesteuerte Software-Evolution. In (Brandt-Pook, H. Hrsg.): Nachhaltiges Software Management. Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformtik (WI-MAW); 14. - 16. November 2012 in Bielefeld. Ges. für Informatik, Bonn, 2012, S. 50–68.

[So11] Sommerville, Ian: Software Engineering. 9th ed., Pearson, Boston, 2011.

[Ti13] Tiemeyer, Ernst (Hrsg.): Handbuch IT-Management: Konzepte, Methoden, Lösungen und Arbeitshilfen für die Praxis. 5.überarb. und erw. Aufl., Hanser, München, 2013.

[Vd11] VDI Zentrum für Ressourceneffizienz GmbH: Umsetzung von Ressourceneffizienz-Maßnahmen in KMU und Ihre Treiber. Berlin, 2011.

[Wo14] Wohlgemuth, Volker; Krehahn, Peter; Ziep, Tobias; Schiemann, Lars: Entwicklung eines Open-Source basierten Baukastens zur Unterstützung und Etablierung der Ressourceneffizienz in produzierenden KMU. In: Wohlgemuth, Volker (Hrsg.) ; Lang, Corinna V. (Hrsg.) ; Marx Gómez, Jorge (Hrsg.): Konzepte, Anwendungen und Entwicklungstendenzen von betrieblichen Umweltinformationssystemen (BUIS). Shaker, Aachen, S. 41 – 57, 2014.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Praxisbeispiel Deutsche Bahn AG: FINK - mit BPM komplexe Fachprozesse im Umweltschutz managen

Svetlana Bloching[1], Jürgen Neumann[2], Iris Rabener[2] und Ingo Rau[2]

Abstract: Die Verschärfung der gesetzlichen Anforderungen an die Verursacher oder Rechtsnachfolger von "Eingriffen" im Sinne des Bundesnaturschutzgesetztes (BNatSchG) vom März 2010 resultierte für die Deutsche Bahn AG in umfangreichen Dokumentations- und Reportingpflichten. Auf Grund des Umfangs, der langen Laufzeiten und sich kontinuierlich verändernder Anforderungen musste ein flexibles und nachhaltiges IT-System geschaffen werden, welches die komplexen Prozesse rund um naturschutzfachliche Kompensationsmaßnahmen aktiv steuert und interne wie externe Anwender bei der umfangreichen Datenerhebung und Berichtserstellung signifikant unterstützt. Mit dem weitestgehend auf Open Source Software basierenden „**F**ach**I**nformationssystem **N**aturschutz und **K**ompensation – FINK“ hat die Deutsche Bahn AG ein umfangreiches Business Process Management System geschaffen, welches die Einhaltung der gesetzlichen Anforderungen gewährleistet und die Erreichung der Konzern eigenen Umweltziele erheblich begünstigt. Wir geben Einblick in den Analyse- und Entscheidungsprozess hin zu Business Process Management und zeigen, wie mittels BPM und Scrum die agile Konzeption, Entwicklung und Einführung von FINK verlaufen ist.

Keywords: BPMN 2.0, Camunda, Liferay, Open Source, Geschäftsprozesse, Agile, Scrum

1 Ausgangssituation und Zielstellung

Die gesetzlichen Anforderungen an die Verursacher oder Rechtsnachfolger von "Eingriffen" im Sinne des Bundesnaturschutzgesetztes (BNatSchG) haben sich seit der Novelle dieses Gesetzes vom März 2010 verschärft [GH10]. Für die Deutsche Bahn AG (DB AG) ergaben sich daraus umfangreiche Dokumentations- und Reportingpflichten. So verlangt das Eisenbahn-Bundesamt (EBA) einen jährlichen revisionsfesten Bericht über die qualitätsgesicherte Umsetzung und dauerhafte Unterhaltung von Kompensationsverpflichtungen.

Zu Kompensationsverpflichtungen kommt es immer dann, wenn Infrastrukturvorhaben nicht vermeidbare Eingriffe in Natur und Landschaft zur Folge haben. Der Verursacher dieser Eingriffe trägt die Verantwortung die Maßnahmen, die dieser Eingriffe kompensieren, umzusetzen und kontinuierlich die Durchführung zu kontrollieren [Bar03, KPW04]. Während der mitunter Jahrzehnte bestehenden Laufzeit übernehmen verschiedene DB Organisationseinheiten die Verantwortung für Herstellung, Ent-

[1] Deutsche Bahn AG, Ressort Technik und Umwelt, Caroline-Michaelis-Str. 5-11, 10115 Berlin, svetlana.bloching@deutschebahn.com
[2] econauten UG (haftungsbeschränkt) & Co. KG, Parkstr. 3, 13187 Berlin, juergen.neumann@econauten.de, iris.rabener@econauten.de, ingo.rau@econauten.de

wicklung und Unterhaltung der Kompensationsmaßnahmen und damit auch die jährliche Berichtspflicht.

Aufgrund der Komplexität, der Laufzeit der rechtlichen Verpflichtungen und der Datenvolumina ist eine anforderungsgerechte Berichterstattung nur mit IT-Unterstützung möglich. Daher wurde im August 2013 mit der Umsetzung des konzernübergreifenden IT-Projekts „FINK“ (**F**ach**I**nformationssystem **N**aturschutz und **K**ompensation) unter fachlicher Schirmherrschaft der DB Umwelt begonnen.

Ziel des Projekts ist es, eine IT-Unterstützung zu konzipieren und einzuführen, die die vollständige Bearbeitung des Gesamtprozesses von der Planung der Infrastrukturmaßnahme bis zum Ende des Pflegezeitraums einer Kompensationsmaßnahme unterstützt und die es DB internen und externen Akteuren ermöglicht, gesetzliche Vorgaben an Dokumentation und Berichtswesen frist- und regelgerecht zu erfüllen. Mit der Einführung des IT-Systems sollen verschiedene wirtschaftliche, organisatorische, technische und qualitative Anforderungen der DB AG erfüllt werden.

Wirtschaftlich gesehen sollen die beteiligten Infrastrukturbereiche der DB nach der Aufnahme des IT-Tools in den Regelbetrieb die unnötigen Kostenrisiken durch mögliche Bußgelder, Kosten für Ersatzvornahme bis hin zur Rückgabe von Bundesmitteln im Zusammenhang mit Kompensationsverpflichtungen vermeiden. Durch optimierte und IT-unterstützte Prozessabläufe werden die Kommunikationsaufwände zwischen den beteiligten Bereichen und den Geschäftsfeldern gesenkt. Der Personalzuwachs für die Bearbeitung der Kompensationsmaßnahmen soll bei der DB Umwelt und den beteiligten Geschäftsfeldern durch die IT-Lösung gering gehalten werden.

Auf der organisatorischen Ebene werden eine durchgängige Bearbeitung sensibler Informationsprozesse und ein einheitlicher Informationsstand für alle am Prozess Beteiligten benötigt. Eine qualitätsgesicherte und effiziente Beherrschung der kontinuierlich wachsenden Datenmengen, Aufgaben und Anforderungen wird angestrebt. Die Geschäftsfelder werden dadurch beim Erreichen der vereinbarten Ziele und der Erfüllung der Legale-Compliance Anforderungen sowie der Verringerung von rechtlichen Risiken unterstützt. Durch schnelle Bereitstellung bedarfsgerechter Produkte (Reports, Gutachten, Umweltkennzahlen etc.) mittels IT-Lösung setzt die DB Umwelt ihre Aufgaben effizienter um.

Das Projekt FINK hat Innovationscharakter, weil keine vorhandene IT-Lösung identifiziert werden konnte, die die Anforderungen erfüllt. Es handelt sich also um eine komplette Neuentwicklung, die mit modernen technischen Standards kompatibel sein soll. Das IT-Tool soll schnell an neue Nutzeranforderungen und mögliche Änderungen der Rahmenbedingungen wie gesetzliche Anforderungen, interne IT-Architektur etc. angepasst werden können. Ziel ist es, die Wartungs- und Weiterentwicklungsmöglichkeiten aller IT-Systeme im Umweltschutz der DB weiter zu optimieren. Eine einheitliche Systemlandschaft über mehrere Geschäftsfelder hinweg mit einer standardisierten prozessorientierten IT-Lösung wird angestrebt. IT-Insellösungen in einzelnen Geschäftsfeldern sollen so vermieden werden

Da das Tool in der Zukunft große Mengen von Umweltdaten verarbeiten soll, sind qualitative Anforderungen an die Datenverarbeitung zu beachten. Die vollständige, qualitätsgesicherte und termingerechte Erfassung von Umweltdaten und entsprechende Lieferung der geforderten Reports innerhalb des DB-Konzerns und an Externe (u.a. Berichte an das Eisenbahn-Bundesamt) sollen mit dem Tool ermöglicht werden. Die Erhebung und Verarbeitung der Daten sollen effizient und nachprüfbar gestaltet werden. Fehler die sich heute aus der manuellen Datenweitergabe ergeben, sollen zukünftig durch die Verringerung von Medienbrüchen und durch qualitätsgesicherte IT-Prozesse vermieden werden.

2 Umsetzungsalternativen

Für die oben beschriebenen DB-spezifischen Anforderungen und Ziele ist keine fertige Lösung am Markt vorhanden. Daher musste die "make-or-buy"-Entscheidung auf "make" hinauslaufen. Für die Entwicklung der Software werden grundsätzlich zwei Umsetzungsalternativen betrachtet: entweder eine flexible Business Process Management Anwendung (BPM), die über Webservice-Schnittstellen mit bestehenden Fachapplikationen verbunden wird, oder eine statisch ausprogrammierte, speziell auf Kompensationsverpflichtungen ausgerichtete eigene Fachapplikation.

2.1 Umsetzungsalternative 1

Entschließt man sich für den Einsatz eines BPM-Systems, können strategische und operative Prozessmodelle rasch und ohne komplexe Übersetzungsschritte in technische Prozessmodelle überführt werden. Einmal aufgesetzt werden im BPM-System Geschäftsprozesse vor allem fachlich modelliert. Das lauffähige Technische Prozessmodell entsteht dann in enger Anlehnung an das fachliche Modell teilweise allein durch Konfiguration, sowie durch weitere spezifische Programmierung. In einem BPM-System können viele verschiedene, voneinander unabhängige Geschäftsprozesse konfiguriert werden. Damit steht ein nachhaltiges System zur Verfügung, mit dem neben dem reinen Berichtswesen zu Kompensationsverpflichtungen schrittweise auch andere Aufgaben im Umweltkontext IT-technisch unterstützt werden [FR2014].

Ein BPM-System kann erheblich iterativer und agiler eingeführt werden, als dies bei klassischen, monolithischen IT-Applikationen möglich ist. Iterativ heißt hier, dass nach dem Aufbau der BPM-Basis zunächst überschaubare Teilprozesse modelliert und technisch abgebildet werden können. Alle Beteiligten – von den Fachverantwortlichen bis zu Prozess- und IT-Spezialisten – gewinnen Erfahrungen, wie das System optimal konfiguriert ist. Statt dem Schreiben vollumfänglicher Fachfeinkonzeptionen vor Beginn der eigentlichen Programmierung (wie bei klassischen Software-Entwicklungsprojekten häufig der Fall) werden Anforderungen rasch umgesetzt und im Testbetrieb angepasst,

bis optimale Ergebnisse erzielt werden. Umfang und Komplexität der abgebildeten Prozesse werden sukzessive gesteigert. Den Prozess tangierende Fachapplikationen und Datenquellen können schrittweise angebunden werden [Lin13].

2.2 Umsetzungsalternative 2

Alternativ zur BPM Lösung kann eine bestehende Fachapplikation im Kontext von Kompensationsverpflichtungen um die benötigten Funktionen erweitert werden, oder eine neue "stand-alone" Fachapplikation entwickelt werden.

Für die Anpassung oder Neuentwicklung einer einzelnen Fachapplikation sollen beim klassischen Vorgehen zunächst alle Anforderungen in einem Fachfeinkonzept beschrieben und anschließend individuell ausprogrammiert werden. Änderungen am operativen Prozessmodell z. B. aufgrund sich ändernder Abläufe im DB Konzern oder sich ändernder rechtlicher Rahmenbedingungen müssen als komplexe Änderungsanforderungen (Change Requests) erneut vollständig spezifiziert und in der Software implementiert werden [HRS14].

3 Bewertung und Empfehlung

Die Analyse des Themenfelds Kompensationsverpflichtungen ergab, dass stark prozessorientiert gearbeitet wird, denn die Aktivitäten, die zu einer Kompensationsverpflichtung führen oder die während der Laufzeit der Kompensationsverpflichtung stattfinden müssen, haben nicht nur eine starke fachliche Dimension (Was ist zu tun?), sondern auch immer eine sachlogische und zeitliche Dimension (In welcher Reihenfolge und wann ist etwas zu tun?).

Die beteiligten Fachverantwortlichen besitzen stark ausgeprägte Fach- und Prozess-Kompetenzen. Es fehlten ihnen jedoch Werkzeuge, um die Prozesse digital abzubilden. Bisher existierten (wenn überhaupt) nur Fachapplikationen, die einen Teilschritt innerhalb des Gesamtprozesses unterstützten. Die Informationsweitergabe zwischen den Systemen erfolgte auf Basis von Dokumenten (digital oder i. d. R. auf Papier).

Von daher erschien die Einführung einer Business Process Management (BPM) Lösung empfehlenswert, um möglichst alle formulierten Projektziele kosteneffizient erreichen zu können. Diese Lösung wird als besonders relevant für die automatische Überwachung von Fristen, Grenzen und anderen Vorgaben durch aktive Systembenachrichtigung sowie effiziente Controlling gesehen. Des Weiteren ermöglicht BPM die Planungssicherheit durch IT-gestützte Prognosen [Eid12]. Auf Basis der im Prozess verfügbaren Daten soll das System in der Lage sein, notwendige Änderungen in der Planung berechnen zu können.

Die Analyse der Anforderungen ergab, dass der Anteil menschlicher Interaktion in den

digital abzubildenden Prozessen auch in Zukunft hoch sein wird. Es geht also nicht primär um Automation oder sogenannte „Dunkelverarbeitung“ bei der ein Prozess komplett eigenständig von der Prozess-Engine abgearbeitet wird. Vielmehr sollen die genannten umweltfachlichen Experten mit dem System interagieren und entlang prozessgesteuerter Vorgaben Entscheidungen fällen. Um das zu erreichen, kommt der Frage nach der Benutzungsoberfläche große Bedeutung zu. Im vorliegenden Fall fiel die Wahl auf das Web-basierte OpenSource Portal Liferay, das sich hervorragend mit der BPM-Lösung Camunda integrieren lässt [Cam15].

Der Einsatz einer Web-basierten BPM Lösung bringt längerfristig Vorteile durch hohe Flexibilität und gute Anpassbarkeit. Er ist sinnvoll, wenn Wachstum geplant ist und der Vernetzungsgrad zwischen Fachabteilungen untereinander und zu externen Beteiligten weiter zunehmen wird [FR14]. DB interne wie auch externe Prozessbeteiligte können über ein einheitliches Webinterface angebunden werden, um ihnen Dateneingabe, Datenmanipulation, Aufgabenmanagement und Datenanalyse zu ermöglichen. Die aufwändige Installation und Wartung der Software auf Client-Computern diverser Anwender entfällt.

Um allen Projektbeteiligten die Lösungsempfehlung nachvollziehbar zu gestalten, wurde eine SWOT-Analyse (siehe Abb.1) für BPM in der IT-Landschaft der DB durchgeführt. Die Analyse bestätigt die Entscheidung für die BPM-Technologie.



Abb. 1: SWOT-Analyse BPM

4 Vorgehen und Projektaufbau

4.1 Konzeption und Umsetzung

Im Rahmen der Vorstudie erwies sich die *Business Process Model and Notation – BPMN 2.0* als besonders gut geeignet die verschiedenen Perspektiven der Prozessbeteiligten zu visualisieren (siehe Abb. 2). Gemeinsam mit den Fachverantwortlichen wurde zunächst ein Soll-Prozess auf hoher Abstraktionsebene entwickelt und ein auf BPM basierendes Lösungskonzept skizziert. Ziel dieser Aktivitäten war es, mit allen Beteiligten über den Prozess in einer einheitlichen, formalisierten Form reden zu können bzw. eine gemeinsame Sprache zu finden, um Rollen, Entscheidungen und Informationsflüsse sichtbar zu machen.

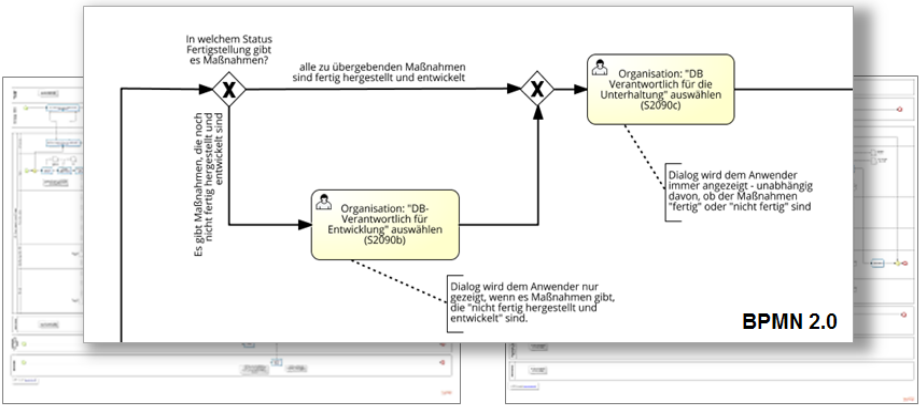


Abb. 2: Beispiel für einen modellierten Fachprozess mit BPMN 2.0

Im nächsten Schritt wurden die gewonnenen Erkenntnisse aufgegriffen und im Rahmen einer Machbarkeitsstudie näher untersucht. Eine Prüfung, ob die vorgeschlagene BPM-Struktur in die bestehende DB IT-Infrastruktur integriert werden kann, war hierfür erforderlich. Detaillierte Informationen zu Systemen wurden beschafft, die mit dem BPM-System integriert werden sollten. Folgende Fragen waren dabei zu beantworten: Welche Integrationsmöglichkeiten existieren? Mit welchem Kosten/Nutzen Verhältnis können Schnittstellen geschaffen werden? Unter welchen Voraussetzung kann auf die Systeme zugegriffen werden (Stichwort: Datenschutz)?

Im Anschluss wurde die Anforderungsbeschreibung weiter vertieft und verschiedenen Nutzungsfälle untersucht, die mit dem System unterstützt werden sollten. Es folgte die ausführliche Analyse der Artefakte und Daten im Prozess – um ein geeignetes Datenmodell und ggfs. notwendige Datentransformationen entwickeln zu können.

Aufwände für die Feinkonzeption (Operatives Prozessmodell) für die technische

Prozessmodellierung und die Unterstützung durch ein externes BPM-Systemhaus zur Umsetzung der zu erwartenden Anforderungen wurden abgeschätzt. Ferner wurden die Möglichkeiten der späteren Anpassung und Weiterentwicklung der Prozesse durch den Auftraggeber analysiert.

Die hier genannten Arbeiten mündeten in einer Entscheidungsvorlage, zu welchen Kosten und in welchem Zeitraum die Einführung einer BPM-Lösung für die Erfassung von Kompensationsverpflichtungen möglich ist. Die in Abbildung 3. dargestellten Meilensteine zeigen die Schritte der Projektdurchführung. Das Projekt wird in folgenden Phasen iterativ realisiert: die Analyse der Daten und Entwurf, die Umsetzung des technischen Modells (Softwareentwicklung und -test) und die fachliche Abnahme und Einführung in den Regelbetrieb.

Ab dem Zeitpunkt, zu dem ein IT-Dienstleister feststeht, bis zu ersten nutzbaren Geschäftsprozessen für die Fachanwender, wurden nur ca. 6 Monate prognostiziert. In den folgenden 2 Jahren wird das System schrittweise bis zur endgültigen Ausbaustufe aufgebaut.

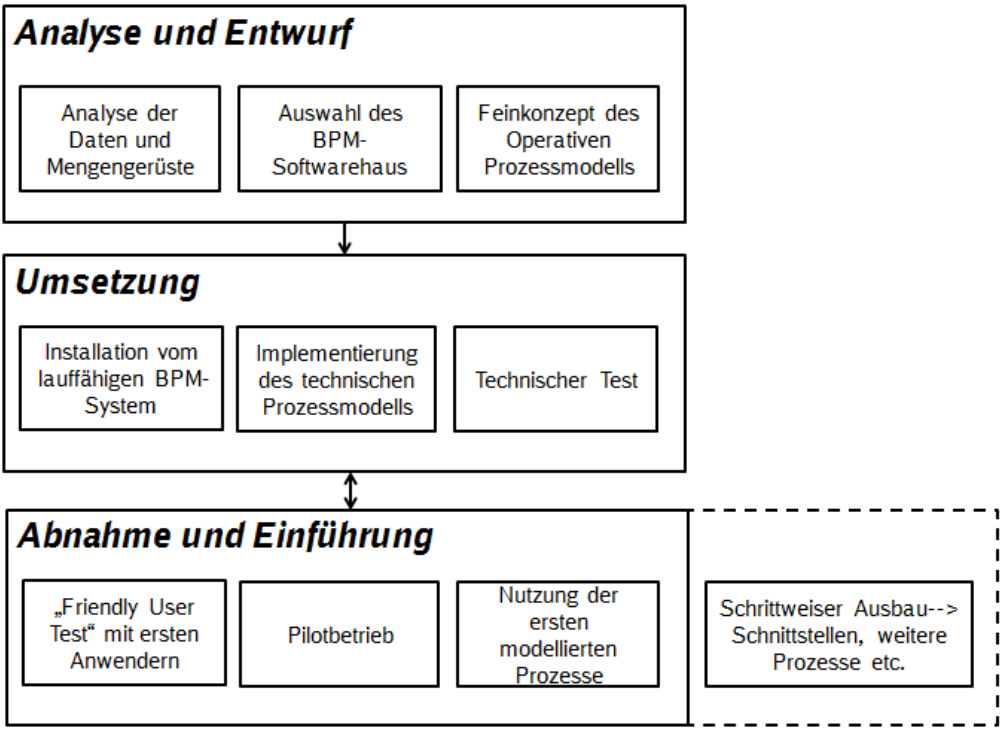


Abb.3: Projektphasen

4.2 Agile Entwicklung mit Scrum

Auf Grund des relativ hohen zeitlichen Drucks erfolgt die Einführung iterativ. Nach Auswahl eines geeigneten BPM-Anbieters auf Basis der Ergebnisse einer Machbarkeitsstudie wird mit der Umsetzung der Anforderungen entlang der agilen Projektmethode Scrum begonnen. In meist 2-wöchigen Entwicklungssprints werden Anforderungen umgesetzt, die zuvor vom Product-Owner priorisiert wurden (siehe Abbildung 4). Projektkoordination und Kontrolle werden vom Auftraggeber DB Umwelt übernommen. Die technische Projektsteuerung sowie die Aufgabe des Scrum-Masters übernimmt eine interne IT-Abteilung der DB AG. Die Vertretung der Anforderungen (Requirements) der verschiedenen Prozessbeteiligten im Projekt-Tagesgeschäft wird in einen so genannten „Product Owner Proxy“ gebündelt. Stellvertretend für DB Umwelt und die anderen DB Geschäftsfelder wird dafür ein auf Softwarekonzeption und agile Vorgehensweisen spezialisiertes Beratungsunternehmen beauftragt. Die Implementierung des IT-Tools übernimmt ein externer IT-Dienstleister, der sich im BPM-Themenfeld etabliert hat.

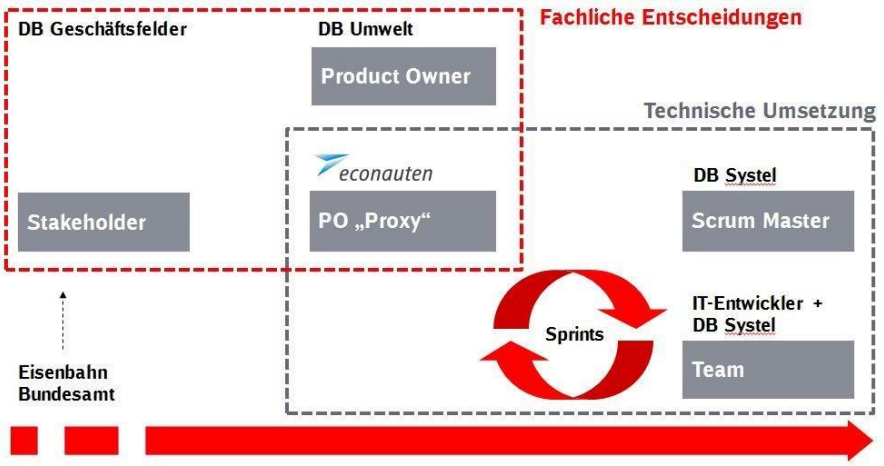


Abb. 4: Projektbeteiligte nach Scrum

5 Open Source Technology Stack

Für das Webframework kamen mehrere Alternativen in Frage. Die Entscheidung fiel schließlich auf Liferay als zugrunde liegende Plattform. Das Open Source Enterprise Portal liefert viele Funktionen einer browserbasierten Business-Software bereits im Standard. Look & Feel, orientiert an DB Styleguide Vorgaben, und eine moderne Benutzeroberfläche gewährleistet das JSF Framework PrimeFaces.

Um die geschäftsfeldübergreifende Arbeit an Kompensationsverpflichtungen möglichst

dicht an den realen Geschäftsprozessen auszurichten, wurde Camunda BPM ausgewählt. Die leichtgewichtige, auf BPMN 2.0 basierende Open Source Plattform eignet sich hervorragend, um Geschäftsprozesse mit BPMN zu modellieren und zu automatisieren [FR14].

Die Entscheidung für ein größtenteils auf Open Source Software basierendes System fiel nicht nur auf Grund der deutlichen Preisunterschiede gegenüber proprietären Alternativen. Relevant war auch, dass Kompatibilität zu der bei DB Systel vorhandenen Betriebsumgebung sowie die einfache Integration des Systems in die sogenannte Standardbetriebsführung der DB Systel sichergestellt werden konnten.

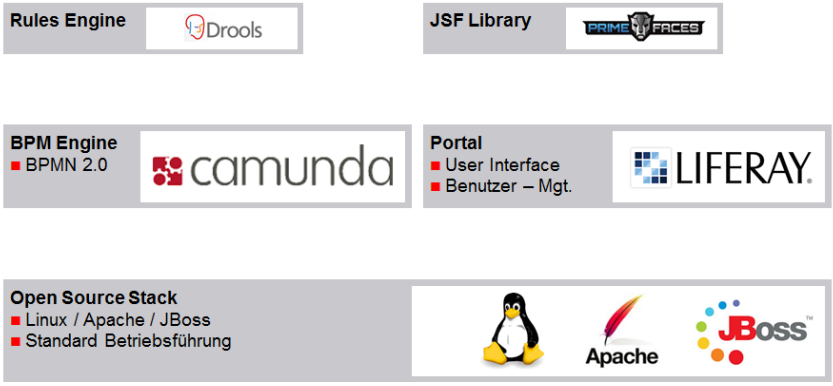


Abb. 5: Verwendete Produkte

6 Fazit

Durch die exzellente Zusammenarbeit zwischen allen Projektbeteiligten und die agile Vorgehensweise war es möglich, FINK in kurzer Zeit für die Pilot-Nutzung freizugeben und in ausgewählten Regionen mit der Korrektur bereits erfasster „Altdaten“ zu beginnen. Schrittweise wird FINK in den nächsten Jahren weiter ausgebaut und beweist schon heute, dass enterprisefähige BPM-Lösungen auf Open Source Basis im DB Konzern möglich sind. Neben der Auswahl geeigneter Technologien wurden Anforderungsmanagement und Entwicklungsmethodik als erfolgskritische Faktoren identifiziert. Es ist zu erwarten, dass sich mit agilen Projektteams und dem hier eingesetzten Technology-Stack zukünftig auch andere innovative IT-Vorhaben im DB-Konzern schnell und qualitätsgesichert umsetzen lassen.

Literaturverzeichnis

[Bar03] Barsch, H.: Eingriffsregelung – Gegenstand und Ziele. In Barsch, H.; Bork, H.R.; Söllner, R. (Hrsg.): Landschaftsplanung – Umweltverträglichkeitsprüfung – Eingriffsregelung. Justus Perthes Verlag, S. 391-392, 2003.

[Cam15] camunda, http://de.slideshare.net/camunda/webinar-camunda-und-liferay, Stand: 01.06.2015.

[Eid12] Eid-Sabbagh, R.H. et. al.: A Platform for Research on Process Model Collections. In (Weidlich, M. Eds.): Proc. 4th Int. Workshop, BPMN 2012, Vienna, Austria, September 2012, Proceedings. Springer, S. 8-21, 2012

[FR14] Freund, J.; Rücker, B.: Praxishandbuch BPMN 2.0. Hanser, 2014.

[GH10] Gassner, E.; Heugel, M.: Das neue Naturschutzrecht. C.H. Beck, 2010.

[HRS14] Heinrich, L.J.; Riedl, R.; Stelzer, D.: Informationsmanagement: Grundlagen, Aufgaben, Methoden. De Gruyter Oldenbourg, 2014.

[KPW04] Köppel, J.; Peters, W.; Wende, W.: Eingriffsregelung Umweltverträglichkeitsprüfung FFH-Verträglichkeitsprüfung. Ulmer, 2004.

[Lin13] Link, M.: Anforderungskonforme Prozessmotivation - Ein stufenweiser Ansatz zum anforderungskonformen Prozessmanagement mit BPMN. Verlag Dr. Kovac, 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Documentational aspects of integrated management systems on the example of a plastics manufacturing company

Richard Rößler[1], Helen Lachmann[2], Werner Esswein[3]

Abstract: Due to increasing competitive pressures and growing and increasing sustainability-oriented stakeholder requirements, companies are forced to shape business processes efficiently by integrating their management systems (MS). Previous research provided more general, theoretically based insights with a lack of practical relevance and links to the theoretical knowledge base. In contrast, this paper provides a comprehensive compilation of current research results that have been obtained on the basis of a literature review, and compares them with the integration processes and problems of a practical case study. The following five specific problem areas were identified: (1) Different objectives, (2) concentration of responsibilities on a few people, (3) limited information value of key figures, (4) low synchronization of documents, and (5) absence of a coherent information database. The documentation and IT support being influenced the most. Findings regarding documentational aspects of integrated management systems (IMS) are discussed in the context of the current state of research. The urgent need for further research efforts in this area is underlined.

Keywords: Integrated Management Systems, Management Systems Documentation, ISO 9001, ISO 14001, ISO 50001, Case Study

1 Introduction

At the present time, companies are in a fierce competition and under constant pressure to achieve cost savings and a lower consumption of resources. Thus, there is a need to continuously adapt changing market and framework conditions. In addition, they have to face diverse stakeholder needs, which have greatly increased in recent years [Ló10]. First of all, the main focus was on the satisfaction of customers. Nowadays this also applies to the expectations of employees, shareholders and society [ZS05]. Due to the complexity of the requirements a systematic approach with clear structures is necessary. As a result, more and more frequently standards-based MS are used, which serve to focus and coordinate the achievement of objectives within an organization [As10a]. These include for example standards for the management of quality (ISO 9001, [In08]),

[1] Technische Universität Dresden, Lehrstuhl für Wirtschaftsinformatik, insb. Systementwicklung, Münchner Platz, 01062 Dresden, richard.roessler@tu-dresden.de
[2] Technische Universität Dresden, Lehrstuhl für Wirtschaftsinformatik, insb. Systementwicklung, Münchner Platz, 01062 Dresden, helen.lachmann@mailbox.tu-dresden.de
[3] Technische Universität Dresden, Lehrstuhl für Wirtschaftsinformatik, insb. Systementwicklung, Münchner Platz, 01062 Dresden, werner.esswein@tu-dresden.de

environment (ISO 14001, [In04]), energy (ISO 50001, [In11]), risk (ISO 31000, [In09]) and occupational safety (OHSAS 18001, [Oh07]), whereby quality and environmental management are the most commonly used MS [Si14].

The application of various separate MS is usually associated with big operating expenses and efficiency losses, for example due to the duplication of tasks and sub-optimal use of resources [ADV14]. These disadvantages can be avoided by companies by integrating their required MS and thereby fulfill all relevant stakeholder requirements in an efficient and effective manner. The objective of an integrated management system (IMS) is to integrate different business processes in a central system and to make processes more efficient [ZS05].

However, a problem is the complexity of the implementation process of an IMS and related efforts. Different areas of the company need to work closely together to synchronize strategies, processes and documents. Despite the number of theoretical approaches and guidelines the implementation of an IMS often proceeds in an unstructured way [JK04]. Due to their limited capacities especially small and medium enterprises are often overwhelmed by the sheer number of requirements and relatively high implementation costs [Sa11].

The present work is to provide practical insights and analyze how the integration of MS occurs in a medium-sized manufacturing company. The focus is primarily on the challenges and problems that may occur during the integration process and the documentational aspects of such an MS. Existing research on IMS and related difficulties was mainly carried out on a theoretical and abstract level. In this paper, however, the operations are thoroughly investigated to identify specific problems by using qualitative analysis methods. Early detection of obstacles thus creates the possibility to early counteract and to develop appropriate solving strategies.

This article consists of six sections. Subsequently, the research objective and the research methods are explained. The contents of the third section constitute the theoretical basis for the topic of IMS. Section four contains the conduction of the case study. In section five the results of the case study are presented and their transferability to other companies is discussed. Furthermore, findings are discussed in the context of the current state of research. The article concludes with a summary of the main findings, a critical appraisal and an outlook on future research directions.

2 Research Design

The aim of this paper is to investigate the subject area of IMS. Based on a case study, the main focus of the analysis is on the barriers and problems in the integration process. A case study is a special form of qualitative-empirical methodology, in which a current phenomenon is investigated in depth and within its real context [Yi09]. Case studies are preferably carried out for the analysis of complex or innovative issues as they allow a

comprehensive and detailed picture of social reality. On the other hand, case studies have openness to new insights and thus promote the development of new fields of research [BG07].

For conducting the case study a personal interview in the form of an expert interview was used as the primary method of data collection, as the technical and background knowledge of a specialist is needed. A document drafted from the findings of the literature review served as an interview guideline. The following research questions should be answered:

RQ1. What specific problems and constraints occurred during the implementation and maintenance of the IMS in the case company?

RQ2. What are the challenges in terms of documentation and how do these findings fit into the state of research?

3 Literature Review

Typically, companies have a wide variety of motivational reasons to integrate their MS. On the one hand, there are external incentives that require the implementation of an IMS. These include, for example, the satisfaction of customer expectations, pressure from the government and competitors or regulations from the parent company [ZTL10]. On the other hand, the independent use of individual function-specific MS in practice may cause disadvantages, which may hinder their internal operations. Amongst others, these include an increased management overhead, the formation of subcultures in the company and thus a slower exchange of information [ZSL07].

By merging the implemented MS and the related use of synergies a sleek, efficient system is created, which can avoid these disadvantages [Sa11]. An analysis of existing literature shows that in addition numerous benefits can be realized by the integration of MS. The possible advantages mainly can be achieved in the following five areas: (1) internal processes (e.g. [ZTL10]), (2) communication and motivation (e.g. [KW98]; [Ló10]), (3) costs (e.g. [ZTL10]), (4) documentation and paper work (e.g. [Ze11]), as well as (5) audit and certification (e.g. [ZS05]) (Figure 1).

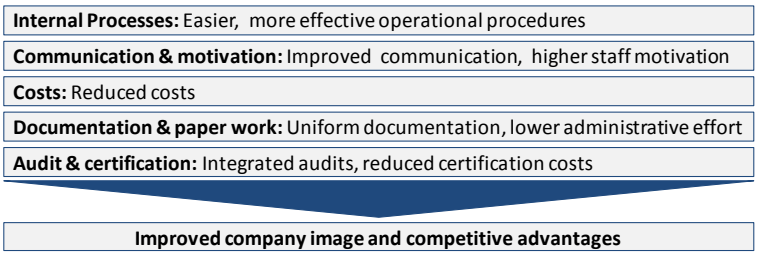


Fig. 1: Advantages of IMS

Due to the different requirements that need to be combined, the integration of MS is a challenging process. In addition, every organization is structured individually, pursues certain goals and has different resources and initial conditions. For this reason, it is not possible to develop a universal solution for an implementation strategy of an IMS [RSS14]. Rather, a general methodology or guideline is needed, which can be tailored to the company's individual operating conditions and objectives [Ló10]. For this purpose numerous theoretical models are presented in literature. Table 1 provides an overview of the main approaches. It should be noted that the approaches are similar to some extent and therefore sometimes it is difficult to distinguish them from each other.

Methodology	**Short Description**
Stepwise approach	Integration can be carried out in a stepwise manner where it proceeds from partial to full integration [Be03]
Integration at various hierarchical levels	Integration needs to cover activities at all hierarchical levels in the organization [JRM06]
Integration through a „TQM"- approach	Use of integrated resources to achieve satisfaction of all stakeholders operating in a TQM [WD01]
Enhancing the MS standards	Integration and enhancing of existing and prospective MS standards [RSK07]
Systems approach to integration	Business is viewed as a single amorphous system that changes its shape depending on prevalent stakeholders [As10a]
Process embedded design of IMS	IMS is designed around the core processes focusing on stakeholder requirements [As09]

Tab. 1: Methodologies for the integration of MS

Besides the advantages, experiences from the introduction of an IMS show that companies also face a variety of problems and obstacles that hinder the integration process [Si14]. Depending on the conditions and characteristics of the organization, these barriers may vary and affect the implementation process in various ways. In analogy to [Be12] integration difficulties can be grouped in three categories: (1) standards-related difficulties, (2) external difficulties as well as (3) internal difficulties (Table 2).

Problem categories	**Possible problems and obstacles**
Standards-related difficulties (e.g. [As09])	• Different scope and structure of the MS • Inadequate standard harmonization • Different models used in the standards
External difficulties (e.g. [RSS14])	• Lack of international standard • Lack of government support • Continuous change of regulations and guidelines

	• Lack of support from certification bodies • Lack of experienced consultants
Internal difficulties [Si14]	• Lack of qualified employees • Low motivation or resistance of employees • Lack of knowledge and understanding for integration • Lack of resources • Lack of strategic planning • Inappropriate organizational culture and structure

Tab. 2: Problems of IMS

4 Case Study

A company from the plastics processing industry is used in order to perform the case study. The corporation represents a suitable case example as it has already implemented several MS, which have been linked into an IMS. According to the case study theory it can be seen as a "typical" case, i.e. the underlying circumstances and conditions of the company examined could also apply in a similar form to other organizations. The company, which has been examined in the case study, is one of the production sites of a group from the field of plastic processing industry and is located in East Saxony/ Germany. Altogether almost 150 employees work at the production site. In 2014 a revenue of about €40 Million was achieved. At the site a total of five certified MS are being operated. They are presented in Table 3 including the year of their first certification.

Management aspect	**Underlying standard**	**First certification**
Quality management	ISO 9001:2008	1993
Hygiene management	HACCP	2002
Environmental management	ISO 14001:2004	2002
Quality management for food packages	BRC/IoP	2005
Energy management	ISO 50001:2011	2013

Tab. 3: Implemented MS of the sample company

The management representative for occupational safety, for environment and energy was chosen as the interview partner for the present case study. He is one of the persons mainly responsible for the integration of MS within the production site. Due to the close cooperation with other management representatives of the group, the interviewee has a

good overview of all procedures within the company. He has been involved in the integration process from the beginning on and therefore fulfills the role of an expert.

An appointment at the locations of the company has been agreed to in order to conduct the interview. The conversation lasted nearly two and a half hours and was based on a previously developed interview guideline[4]. Along with the answering of the questions an insight into the documentation- and IT-structures of the company was granted.

5 Results

5.1 Analysis of the case study

The analysis of the interview reveals that some of the theoretical approaches from the literature review also occur in the case example. They include, for example, customer requirements and guidelines on the part of the parent company as external incentives as well as reduction of costs as an internal incentive. Regarding the implementation strategy the company primarily build a quality management system according to ISO 9001 and integrated the other MS afterwards. This corresponds to the first integration strategy by [KW98]. The application of a special integration methodology, however, could not be determined. The implementation of a new MS is rather based on the associated standard.

According to the interview analysis, basically five main problems can be distinguished (Figure 2). These include *differences in the objectives* between the parent and the production plants. Due to this divergences conflicting goals can emerge which need to be determined and adjusted. Another problem is the *concentration of responsibilities* on only a few people. On the one hand, this leads to a concentration of knowledge on individual employees. On the other hand, the number and variety of tasks may cause the result that the employees responsible have less time for developing the system. A third important aspect is the *limited information value of the key figures used*. Thus, an evaluation and comparison of the production sites is difficult to achieve. Furthermore, the *low synchronization of documents* in the different locations leads to a very extensive thus confusing documentation and to duplication of records. This causes redundancies, which complicate the daily operational processes. The last problem area is the *absence of a coherent information database*. As a result, the maintenance and updating of documents is very work-intensive and inconsistencies between documents and systems occur. Moreover, it is difficult to create a comprehensive overview of relevant group processes.

[4] Similar to [EH06] the following six question areas were developed on the basis of the literature review: (1) motivation, (2) planning of integration, (3) implementation of integration, (4) integration level, (5) obstacles and problems as well as (6) advantages and disadvantages of the IMS. The questionnaire can be requested from the authors.

As already mentioned, the problems result primarily from the different framework conditions at the parent and production plants. These include, in particular, the differing product portfolios as well as the variation in the implemented MS. Due to these divergences, the systems have to be aligned constantly and an easy transfer of the elements is not possible. These interface problems have to be recognized and resolved through the development of the IMS.

The results of the case study have also shown that the theoretical concepts from the literature are partially inconvertible or even occur in another form. For example, the appointment of a single integration representative for the entire IMS is not possible as the required expertise and effort would be too high. A similar problem arises for the implementation of integrated audits.

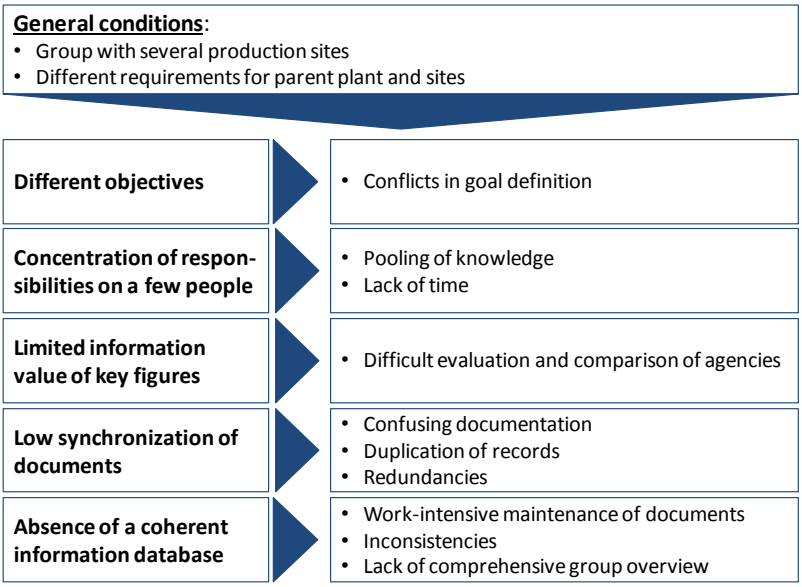


Fig. 2: Identified problem areas from the case study

There also exist differences between theory and practice related to the corporate problems and obstacles. For example, the external difficulties mentioned in the literature review were not addressed during the interview. A similar fact was noticed with the standards-related difficulties, which were not mentioned by the management representative either. In this context it was only stated that the practical relevance of some requirements, which derived from the management standards, could not be clearly determined. The request for more practically oriented standards becomes apparent. One aspect, which could be found in theory and at the company, is the lack of resources, particularly the lack of temporal resources. Furthermore, also the prevailing group structure plays a key role. Particular obstacles for the comprehensive implementation of an IMS are the different structures and strategic focuses of the several group locations.

It is anticipated that the five identified problem categories from the case study may also occur in other companies. This would primarily be possible if similar circumstances in the sample company apply, e.g. in a group with several production sites where different framework conditions occur in the parent factory and the agencies.

5.2 Documentational aspects of IMS

From the results of the case study the particular relevance of the information and documentation aspect can be derived. Three of the identified five problem categories concern this aspect either directly or at least indirectly. While *Low synchronization of documents* and *Absence of a coherent information database* directly indicate problems of an inconsistent and partially redundant documentation, *Concentration of responsibilities on a few people* rather addresses the aspect of knowledge concentration on individuals and an impaired availability of information to third parties. These findings are in line with the findings in literature.

Various authors confirm that the documentation can have a significant impact on the overall success of an IMS. According to Simon et al. [SKC12], an insufficiently integrated documentation is the biggest obstacle to proper management. Pho and Tambo [PT14] emphasize issues related to the document flow and the complex adaptation of function-specific aspects to new documentation requirements of an IMS. Von Ahsen [Vo14] actually concludes that that the cost reduction derived from the introduction of an IMS can be eliminated by a very complex documentation.

To address the challenges of an integrated management, several solutions were discussed in literature, addressing the issue of documentation more or less. Abad et al. [ADV14] and de Oliveira [De13] emphasize that an effective integration of MS standards can only be achieved through an intensive documents integration, which forms the basis for a robust IMS. In this context the company needs to build an efficient document management [SKC12]. Pho and Tambo [PT14] emphasize that for effective document control electronic data management solutions need to be used. In the synergetic multi-level model of Zeng et al. [ZSL07] document integration is the product of an efficient integration of strategic aspects pursued by the corporate management. Asif et al. [As10b] focus on the aspect of the integration of documents. They propose a common strategy in an IMS manual, which forms the basis for the further integration process. Griffith and Bhutto [GB09] recommend the users of their proposed framework to support the introduction of an IMS by information technology. This is to promote an intranet-based documentation. The procedure of López-Fresno [Ló10] includes the analysis of the document structure in the first step. They recommend to design an integrated documentation based on a requirements matrix.

The common feature of these approaches is that the issues of documentation are rather discussed only in passing. The corresponding statements are more like a simple guidance and provide less tangible support to build an integrated documentation. According to the authors' knowledge, there are currently only two studies, which bring the aspects of

documentation into the focus of attention.

Ibarra and Calarge [IC14] recognize the need to define an information structure in an organized and efficient way to improve the performance of the information systems. As part of their work, they propose the Zachman Framework for the definition of an enterprise architecture and adapt it to the application of the integration of MS standards. The definition of enterprise architecture through this framework is the basis for the development of a corporate information system. However, the authors present the considerations rather in the form of a thinking stimulation than as a concrete artifact, so that the opportunity to direct applicability is not given. Roessler and Schlieter [RS15] develop a model-based method for establishing and maintaining an IMS. Based on a content and structure analysis of the requirements of ISO 9001, ISO 14001, ISO 50001 and OHSAS 18001 they identify four groups of requirement types. Accordingly, requirements may be identical, integrated, parallel or different. Based on these so-called integration types design templates are proposed to solve the problems of integration within a meta-model, which forms the basis for integration.

It needs to be stated that the issue of integrated documentation is yet often neglected and addressed only by relatively non-specific recommendations. Apart from the works of Roessler and Schlieter [RS15] and Ibarra and Calarge [IC14], there are no studies that address the issues of documentation as an explicit problem of IS research. The need for the continuation of the previous considerations is thus given and represents a wide field for further research.

6 Conclusions

The examination of the production company from the case study revealed that the theoretical findings of the literature review could only partly be confirmed. During the expert interview some of the internal and external motivations as well as advantages quoted in the literature were also identified in the examined case.

Moreover, some interface problems were recognized which are due to the different structure of the parent plant and the production site and which are not mentioned in the literature. In total, five problem categories were deduced from the case study, which could also occur in companies with similar framework conditions. These include (1) different objectives, (2) concentration of responsibilities on a few people, (3) limited information value of key figures (4) low synchronization of documents, and (5) absence of a coherent information database. Documentation and IT support are regarded as the most influenced internal sectors. In contrast, the staff of the company, which was mentioned as the major obstacle in the literature, plays only a subordinate role in the case example. Here, the differences between theory and practice become obvious.

In the discussion of the case study results it became clear that in particular the aspect of

the documentation and maintenance of the IMS is of particular relevance for the company. It was shown that the insights gained in this regard seamlessly fit into the existing state of research. Addressing the issues associated with the described documentational aspects open up a wide field for further research activities.

The findings of the case study are limited as the applied research methods also show some weaknesses besides their particular strengths. One limitation appears from the implementation of a single-case study making a comparison with other companies impossible. The questionnaire developed as part of this case study can be used as a tool for more extensive field studies. These could provide, for example, interesting insights on the specific characteristics of integration with respect to different company sizes and types. Furthermore, the theoretical findings gained from the literature review could be empirically tested in a broader way. However, the consideration of a single case allows a more intensive analysis of both, the internal structures as well as the cause-and-effect relationships, compared to a multiple case study. To validate the results of this article, additional case studies on this topic should be conducted. In addition, it has to be examined if further problem categories can be identified in other companies with similar conditions.

It becomes clear that the present article opens up new areas of research, which should be examined in future studies. Previous findings provide a platform for further research, especially in the field of IMS in small and medium production companies.

References

[ADV14] Abad, J.; Dalmau, I.; Vilajosana, J.: Taxonomic proposal for integration levels of management systems based on empirical evidence and derived corporate benefits. Journal of Cleaner Production 78(18), pp. 164-173, 2014.

[As09] Asif, M.; de Bruijn, E. J.; Fisscher, O. A. M.; Searcy, C.; Steenhuis, H.: Process embedded design of integrated management systems. International Journal of Quality & Reliability Management 26(3), pp. 261-282, 2009.

[As10a] Asif, M.; Fisscher, O. A. M.; De Bruijn, E. J.; Pagell, M.: An examination of strategies employed for the integration of management systems. The TQM Journal 22(6), pp. 648-669, 2010.

[As10b] Asif, M.; De Brujin, E. J.; Fisscher, O. A. M.; Searcy, C.: Meta-management of integration of management systems. The TQM Journal 22(10), pp. 570-582, 2010.

[Be03] Beckmerhagen, I. A.; Berg, H. P.; Karapetrovic, S. V.; Willborn, W. O.: Integration of management systems: focus on safety in the nuclear industry. International Journal of Quality & Reliability Management 20(2), pp. 210-228, 2003.

[Be12] Bernardo, M.; Casadesus, M.; Karapetrovic, S.; Heras, I.: Do integration difficulties influence management system integration levels? Journal of Cleaner Production 21(1), pp. 23-33, 2012.

[BG07] Borchardt, A.; Göthlich, S. E., 2007. Erkenntnisgewinnung durch Fallstudien (engl.: Gaining knowledge through case studies). In: (Albers, S.; Klapper, D.; Konradt, U.; Walter, A.; Wolf, J. Eds.): Methodik der empirischen Forschung (engl.: Methodology of empirical research), 2nd ed., Gabler, Wiesbaden, 2007.

[De13] De Oliveira, O. J.: Guidelines for the integration of certifiable management systems in industrial companies. Journal of Cleaner Production 57, pp. 124-133, 2013.

[EH06] Eriksson, H.; Hansson, J.: Integrated Management Systems - Theoretical and Practical Implications. The Asian Journal on Quality 7(2), pp. 69-82, 2006.

[GB09] Griffith, A.; Bhutto, K.: Better environmental performance. Management of Environmental Quality 20(5), pp. 566-580, 2009.

[IC14] Ibarra, L. A. C.; Calarge, F. A.: A Proposal of leverage improvements on Integrated Management Systems applying Zachman's Corporative Architecture. Advanced Materials Research 860-863, pp. 3035-3040, 2014.

[In08] International Organization for Standardization: ISO 9001: Quality management systems – Requirements. Geneva, 2008.

[In04] International Organization for Standardization: ISO 14001: Environmental management systems – Requirements with guidance for use. Geneva, 2004.

[In09] International Organization for Standardization: ISO 31000: Risk management – Principles and guidelines. Geneva, 2009.

[In11] International Organization for Standardization: ISO 50001: Energy management systems – Requirements with guidance for use. Geneva, 2011.

[JK04] Jonker, J.; Karapetrovic, S.: Systems thinking for the integration of management systems. Business Process Management Journal 10(6), pp. 608-615, 2004.

[JRM06] Jørgensen, T. H.; Remmen, A.; Mellado, M. D.: Integrated management systems – three different levels of integration. Journal of Cleaner Production 14(8), pp. 713-722, 2006.

[KW98] Karapetrovic, S.; Willborn, W.: Integration of quality and environmental management systems. The TQM Magazine 10(3), pp. 204-213, 1998.

[Ló10] López-Fresno, P.: Implementation of an integrated management system in an airline: a case study. The TQM Journal 22(6), pp. 629-647, 2010.

[Oh07] OHSAS Project Group: OHSAS 18001: Occupational health and safety management systems – Requirements. London, 2007.

[PT14] Pho, H. T.; Tambo, T.: Integrated management systems and workflow-based electronic document management: An Empirical study. Journal of Industrial Engineering and Management 7(1), pp. 194-217, 2014.

[RSS14] Rebelo, M. F.; Santos, G.; Silva, R.: A generic model for integration of Quality, Environment and Safety Management Systems. The TQM Journal 26(2), pp. 143-159, 2014.

[RSK07] Rocha, M.; Searcy, C.; Karapetrovic, S.: Integrating Sustainable Development into Existing Management Systems. Total Quality Management 18(1-2), pp. 83-92, 2007.

[RS15] Roessler, R.; Schlieter, H.: Towards Model-based Integration of Management Systems. In (Thomas, O.; Teuteberg, F. Eds.): Proceedings of the 12th International Conference on Wirtschaftsinformatik (WI 2015), Osnabrück, pp. 31-45, 2015.

[Sa11] Santos, G.; Mendes, F.; Barbosa, J.; Willborn, W. O.: Certification and integration of management systems: the experience of Portuguese small and medium enterprises. Journal of Cleaner Production 19(17), pp. 1965-1974, 2011.

[SKC12] Simon, A.; Karapetrovic, S.; Casadeus, M.: Difficulties and benefits of integrated management systems. Industrial Management & Data Systems 112(5), pp. 828-846, 2012.

[Si14] Simon, A.; Yaya, L. H. P.; Karapetrovic, S.; Casadesus, M.: Can integration difficulties affect innovation and satisfaction?. Industrial Management & Data Systems 114(2), pp. 183-202, 2014.

[Vo14] Von Ahsen, A.: The Integration of Quality, Environment and Health and Safety Management by Car Manufacturers - A Long-Term Empirical Study. Business Strategy and the Environment 23(6), pp. 395-416, 2014.

[WD01] Wilkinson, G.; Dale, B. G.: Integrated management systems: a model based on a total quality approach. Managing Service Quality 11(5), pp. 318-330, 2001.

[Yi09] Yin, R. K.: Case Study Research: Design and methods. 4th ed., Sage Publications, Los Angeles, 2009.

[ZSL07] Zeng, S. X.; Shi, J. J.; Lou, G. X.: A synergetic model for implementing an integrated management system: an empirical study in China. Journal of Cleaner Production 15(18), pp. 1760-1776, 2007.

[ZTL10] Zeng, S. X.; Tam, V. W. Y.; Le, K. N.: Towards Effectiveness of Integrated Management Systems for Enterprises. Engineering Economics. 21(2), pp. 171-179, 2010.

[Ze11] Zeng, S. X.; Xie, X. M.; Tam, C. M.; Shen, L. Y.: An empirical examination of benefits from implementing integrated management systems (IMS). Total Quality Management 22(2), pp. 173-186, 2011.

[ZS05] Zutshi, A.; Sohal, A.: Integrated management system: The experiences of three Australian organisations. Journal of Manufacturing Technology Management 16(2), pp. 211-232, 2005.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

A Monitoring Framework for Guidance and Risk Control Assistance of Environmental Compliance Officers

Heiko Thimm[1]

Abstract: It is the ultimate goal of environmental compliance management to assure corporate compliance with given regulations. The work processes that are required to fulfill and maintain compliance are usually complex and long running processes that are composed of many interdependent human lead activities. Errors of human work activities and also material and equipment defects can lead to non-compliance. A monitoring framework is proposed that based on data entries of compliance officers can track compliance work processes and provide compliance officers with activity guidance. Main activity failures can be detected through failure detection functions that analyze the activity log, failure indicators, and also external data. Based on the proposed framework environmental information systems can be extended towards actively responding systems that guide compliance officers and provide them with assistance for risk control tasks.

Keywords: Environmental Information Systems, Environmental Compliance Management, Failure Estimation, Risk Assessment, Activity Monitoring, Reactive Information Systems.

1 Introduction

The work area of corporate environmental compliance management is directed at the fulfillment of environmental protection regulations by the company [Gu11] [Im12]. To this end specialists from disciplines such as occupational safety, hazardous material management, fire protection, and transportation safety work together with colleagues from other corporate areas. In general, environmental regulations can be directed at almost all corporate areas [HMB11]. Because of their potential to harm the environment often the areas of product engineering, manufacturing, logistics, plant maintenance and quality management are targeted by environmental regulations. Companies need to be prepared to carry out compliance enforcement activities at different levels of all business aspects. This includes for example activities focused on product properties, production and logistic processes, corporate infrastructure components, and also activities that concern the knowledge and skill level of the workforce. The company faces sanctions ranging from fines, withdrawal of licenses and even permits to mandatory closures and shutdowns when full compliance is not maintained at all times.

The successful completion of compliance management obligations that will lead to environmental compliance largely depends on carefully chosen and managed human-

[1] Pforzheim University, School of Engineering, Tiefenbronner Str. 65, 75175 Pforzheim, heiko.thimm@hs-pforzheim.de

lead enforcement activities [Im95]. However, for human work in general one needs to assume that errors can occur and that work results may not meet given outcome constraints. This rule that human work bears the risk of errors, of course, also applies to compliance enforcement activities. Activity errors such as wrong decisions, wrong judgements, and wrong activity plans can result into non-compliance. In the light of the already enormous amount of legal regulations that is still increasingly growing in the future [BM09] errors of compliance management teams are however most likely to occur. In addition to humans also other aspects such as material errors and equipment defects can cause compliance management activities to fail.

In order to prevent that this can happen compliance officers can be provided with system generated guidance for activity completion and assistance for risk control measures. What is needed as foundation for an approach of a corresponding system with realtime responsive functionality is a component that monitors the human activities and performs checks in order to detect activity errors. A corresponding monitoring framework is proposed in this article. Based on data entries from the human actors of compliance activities the ongoing compliance work processes are tracked and also guided with respect to temporal constraints. Through corresponding failure detection functions that analyze the tracking data, failure indicators, and also company external data activity problems can be detected. In a next step the monitoring framework will be implemented within the environmental compliance management information system CCPro [Th15a] in order to obtain first evaluation results.

The use of information system approaches to enforce environmental compliance regulations has been studied so far by only a few research groups. Freundlieb and Teuteberg [FT09] proposed an approach that targets corporate environmental compliances through data warehousing based advanced reporting and analysis capabilities. A public web-based system which based on user input and an XML regulation framework generates a compliance decision has been developed by researchers of the Stanford University [Ke03]. Researchers from IBM and the University of Stuttgart proposed a system where similar to our approach compliance management activities are logged in database tables in order to detect anomalies that may indicate compliance violations [Ag06].

The remainder of this article is organized as follows. The following Section 2 gives an overview of corporate environmental compliance management obligations and especially describes two core work processes and their activities in more detail. The section also describes through a conceptual data model the main concepts of compliance management processes. Section 3 describes the motivation for system-based provisioning of activity guidance and risk control assistance. Furthermore, the three main activity failures of the compliance management processes that are focused in our work are identified. The proposed monitoring framework is described in Section 4 and concluding remarks are given in Section 5.

2 Environmental Compliance Management Processes

In general a company's environmental compliance management efforts are targeted on three goals [Gu11][HMB11]. First, the organization needs to ensure that at all times it knows all the relevant legal regulations for environmental protection. Second, for all relevant regulations the organization has to determine and implement measures that are required in order to fulfill given requirements. Third, the organization needs to document all considerations, measures, special permits obtained from regulation enforcement agencies, and measures targeted on environmental protection. This documentation task has to be completed in a way that enables the organization at any time to proof the fulfillment of the two previously described obligations. Especially, when auditors from accredited independent agencies conduct an environmental audit they expect the company to perform sufficient proofs of fulfillment. This usually involves the demonstrating of the environmental compliance management work practice, consistency checks of the required documents, and data validation steps with available information systems and tools.

In order to meet these goals companies are advised to define for the compliance management officers corresponding operational level business processes that they need to follow [Im95]. Similar to other corporate work areas which are also governed by a relatively large body of legal and other regulations (e.g., human resource management, financial accounting) one can describe a set of core compliance management processes that are applicable to many companies [BM09] [Im12]. For example, from an earlier empirical investigation [Th15a] of the compliance management work practice we obtained two core data-driven processes that serve as reference processes for the research described in this article. These two processes and their main component activities that need to meet temporal and outcome constraints are:

- The process of recognizing new regulations and of treating new regulations in order to enforce environmental compliance referred to in the following by *New Regulation Management process* or in short *NRM process*. This process can be roughly decomposed into the following sequence of activities (called NRM activities) that are to be documented in terms of their progress, actors, and other aspects:
    1. recognize new regulation and capture regulation in the corporate compliance management register
    2. evaluate relevance of new regulation at the company level, the level of organization units, and the level of company locations
    3. determine for relevant regulations those individuals and organizational units which need to decide about measures to respond to the new regulation so that non-compliance will be avoided
    4. implement the measures
    5. check effectiveness of measures

- The process of recognizing revisions on existing regulations and of treating the revisions in order to enforce environmental compliance referred to in the following by *Regulation Change Management process* or in short *RCM process*. The decomposition of this process yields a similar sequence of activities (RCM activities) as described above.

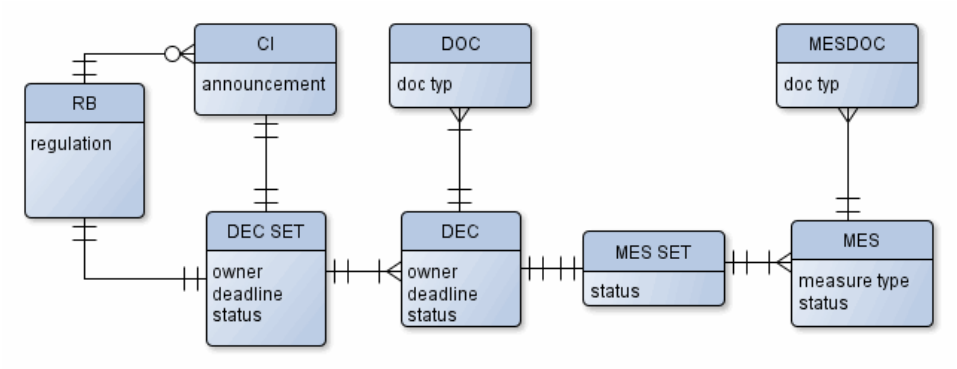


Fig. 1: Conceptual data model for environmental compliance management tasks.

The data objects that are addressed in these two core processes can be described as a corresponding conceptual data model such as given in Figure 1 which uses the popular Martin Notation for cardinalities [Ma90] that is a refined version of the Information Engineering Notation. The definitions of the concepts that capture the meaning of the real world entities addressed in the processes (i.e. object types) are as follows:

- Regulatory Basis (RB): These data objects refer to any kind of regulation that is part of the current general body of environmental laws and that is relevant to the company and thus captured in the regulation register. Every RB object is associated with decisions about compliance enforcement measures.
- Change Incidents (CI): An object of this type represents an announcement of a revision of a relevant regulation or a repeal of a regulation. Thus, every CI object is linked to a particular RB object. Furthermore, CI objects are linked to decisions about compliance enforcement measures.
- Decision sets (DEC SET) and (composite) decisions (DEC): In most cases when compliance enforcement measures are required several alternative actions are existing. This situation leads to a selection decision and this is addressed by a data object type DEC SET. This type represents a set of decisions that are all targeted either at the same CI object or the same RB object. The (composite) decisions that are part of a decision set are represented by the type DEC. It is the goal of each composite decision to obtain a set of proper compliance enforcement measures (MES SET).

- Documents involved in the decision making processes (DOC): For reasons including traceability and transparency the decision making processes are to be documented through a set of documents. The DOC objects are linked to the respective (composite) decision objects.
- Sets of measures (MES SET) and (composite) measures (MES): The data object type MES SET refers to sets of compliance enforcement measures that are to be completed at a certain deadline in order to maintain a given positive environmental compliance status. The (composite) measures that belong to a set of measures are represented by the object type MES. Each measure is of a certain measure type.
- Documents that describe measures (MESDOC): The measures are to be documented with respect to many aspects such as the completion state and the measure effectives through a set of documents. The MESDOC objects are linked to the referring MES object.

3 Motivation for Guidance and Assistance for Risk Control

As of today, environmental law overall is comprised of more than 9000 acts and subordinated regulations [BM09] [HMB11]. The assumption is that the density of environmental law provisions will continue to grow in the future. For years, new laws, directives or amendments to existing laws or previous amendments are being announced. Notices of the repeal of laws (i.e. repeal notices) are no longer the exception. The density of provisions and the amendment dynamic presents organizations with a challenge in keeping up with current, more specifically relevant legislation. But this is exactly what companies are obligated to do in order to achieve and maintain a positive environmental compliance status. Companies are obligated to know and comply with the relevant laws and regulations (obligation to stay informed). Should a company not comply with this obligation (knowingly or unknowingly), the company faces sanctions ranging from fines, withdrawal of licenses and permits to mandatory closures and shutdowns [Gu11]. Naturally, the principle *ignorantialegis non excusat* (ignorance of the law excuses no one) applies. Accordingly, companies hire internal and external environmental law specialists to assure that companies comply with all relevant statutory regulations. The growing complexity however is increasing the probability that compliance experts are over-challenged and thus fail to fulfill required compliance enforcement tasks which will cause corporate non-compliance. In addition to that there always exist the human factors such as compliance officers with mental problems as a further source of corporate non-compliance.

Apart from compliance officers that as individuals can be over-challenged by the sheer complexity of the regulation situation there exist also further sources of the risk of non-compliance. There are first of all the known phenomena of group work such as a destructive group atmosphere and the Group Think effects [Ja72] that can contribute to an increasing risk of non-compliance. In addition to that, measures that are targeted on

maintaining compliance can suffer from material errors and equipment defects. These aspects need to be treated as a further potential cause of the risk of non-compliance.

The potential for the risk of non-compliance motivates in the light of the severe consequences of compliance violations the specific research reported in this article. It is the goal to develop and study information system approaches that, firstly, offer to compliance experts helpful activity guidance [Bo09]. This guidance is directed at the pro-active enforcement of fulfilling both temporal constraints and outcome constraints of activities required by NRM processes and RCM processes. To this end the guidance is especially oriented at proposing appropriate activity time schedules to compliance management experts. Secondly, it is intended to offer to the compliance management experts assistance for risk control tasks directed at the risk of non-compliance. At the current state the focus of the latter is to detect potential risk sources and to present to compliance management experts the risk of non-compliance in the form of a risk diagram [Th15b].

The detection of risk sources is focused on evaluating NRM activities and RCM activities in terms of possible violations of the activities' temporal constraints and outcome constraints. These violations are generally referred to as *activity failures*. In the remainder of this chapter three specific activity failures are described which are referred to by *missed activities, overdue activities,* and *imperfect activities*. Also described are examples for approaches to detect such failures through the evaluation of easy to compute "hard indicators" and also more sophisticated "soft indicators" that can involve company external data and complex failure estimation methods.

By *missed activities* it is referred to activities that at a given point in time need to be executed in order to maintain compliance but however are missing in the set of ongoing activities. A missed activity, for example, can refer to a situation where the compliance team over-looked the announcement of a new community regulation for noise immission that will become effective at a certain future date. Consequently, the new regulation will not be addressed and cause a future non-compliance state. Checking "hard indicators" in order to detect missed activities means to compare the usual pattern of monitoring activities with the most recent pattern. Deviations between the patterns are to be treated as potential activity failures. Checking "soft indicators" can be performed by evaluating external information sources that store regulation announcements such as online information services and regulation portals. For example, assume that one identified a time period in which many new regulation announcements were published. When during this time period only a small number of monitoring activities is found in the corporate compliance management data it can be assumed that the company missed some monitoring activities.

By *overdue activities* it is referred to activities that are either overdue at the present time point or that will miss their activity deadline during a near future time interval. To give an example assume that the above described noise regulation was recognized and in a next step evaluated to be relevant for the company. As further activity it will be required

to decide about measures to fulfill the noise immission regulation. That is, one needs to schedule a corresponding decision activity and set an activity deadline. When the decision activity is not completed in time then it can happen that the measure implementation activity will not be completed in time, too. This activity failure can ultimately cause the company to fail the new noise immission regulation and result into a non-compliance state. In order to detect overdue activities one can simply check for activities that violated their deadlines. In order to complement this approach by an analysis that is directed at "soft indicators" one can check for activities that did not violate a deadline (yet) but that are likely to fail a required deadline. The identification of activities that are likely to fail deadlines can involve many aspects such as the activity complexity and the resources required. For example, consider a complex measure implementation activity that is behind its schedule. When the remaining time span to the deadline is relatively short a relatively high likelihood for a corresponding deadline violation can be assumed.

By *imperfect activities* it is referred to either already terminated activities that violated outcome constraints or still ongoing activities that will terminate during the near future time interval with outcome constraint violations, too. As an example, consider again the scenario of the noise immission constraint. Assume that in a further step it was decided to implement a noise immission fence. An imperfect activity refers to the situation that the fence (i.e. measure being implemented through a corresponding implementation activity) will not well enough shield the noise immission. As a result of the failed outcome constraint the noise limit will be violated which leads to a non-compliance status. The detection of terminated activities that violated outcome constraints is a relatively simple checking task. The detection of constraint violations for still ongoing activities is a more complex task that can be performed, for example, by the use of appropriate prediction methods (e.g. [SLM10]) and by expert judgements.

4 Monitoring Framework

A number of information systems have been developed by research teams and software companies that are in general oriented at corporate environmental management tasks – referred to by *Corporate Environmental Management Information Systems (EM-IS)* [TS09] [Me10] - or that are focused on the corporate environmental compliance management tasks – referred to by *Corporate Environmental Compliance Management Information Systems (ECM-IS)* [BM09]. It appears that in these approaches it is the predominant role of the information system to record activities, document activities, and to analyze and provide information for environmental management activities such as environmental reports. That is, similar to many other corporate information systems the role of EM-IS and ECM-IS has so far been viewed mostly as a passive role. The monitoring framework proposed in this article targets an extension of these systems towards active and self-responsive systems. It is the intension to enable these systems to automatically check conditions and to self-reactively provide online-response that is

helpful for corporate environmental compliance managers. Especially, active guidance and risk control assistance should be provided to the corporate compliance management team [Bo09].

The central pillars of the proposed monitoring framework are two data structures referred to by *Announcement Checking Schedule (ACS)* and *Execution Control Record (ECR)*. These data structures are specialized to the purpose of tracking and monitoring of environmental compliance management activities and the detection of activity failures in realtime as described in Section 3. Figure 2 gives an overview of the main principles of the framework by showing a corresponding concept diagram. The diagram focusses on the activities of the New Regulation Management (NRM) process that are depicted as boxes with activity names as labels. The principles explained in the following are also applicable to the activities of the Regulation Change Management (RCM) process. The functions to detect activity failures based on ACS data and ICR data are depicted as boxes with a gear wheel. The rounded boxes of Figure 2 correspond to data structures.

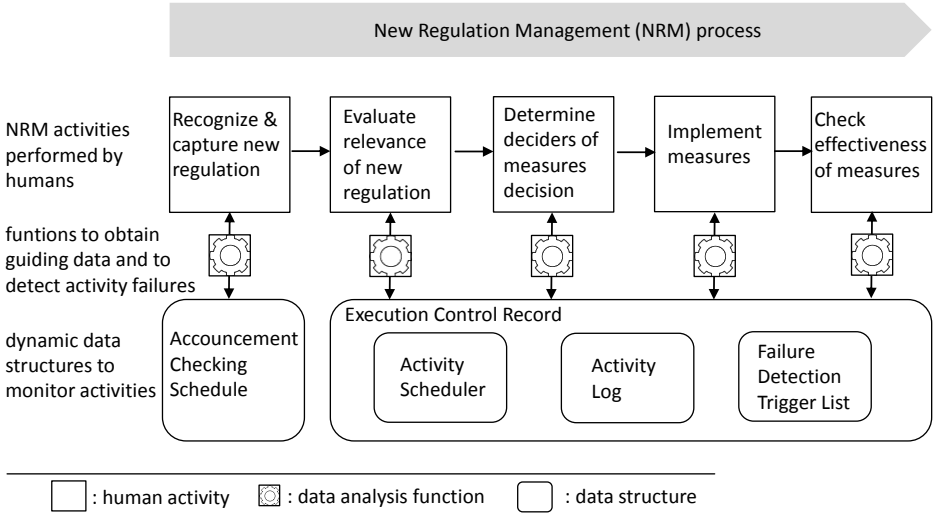


Fig. 2: Main principles of proposed monitoring framework.

The ACS contains both the schedule of required (i.e. planned) activities to monitor and to record new regulations and the set of confirmed (i.e. completed) activities. Based on the activity information it is possible to detect missed monitoring activities.

The ECR data structure consists of a set of three data objects that are created for every new regulation that is inserted in the regulation register. In order to perform an automated monitoring and evaluation of activities in realtime the data objects are dynamically revised and extended. The revision and extension operations are completed during the execution of the activities with the goal to accumulate data for detecting activity failures.

The *Activity Scheduler* specifies deadlines and time buffers for the activities that belong to NRM processes and RCM processes. The *Activity Log* records details about the activity execution. The *Failure Detection Trigger List* maintains a list of predefined triggers that enact failure detection functions. As described earlier on the basis of the data available in the ECR data structure it is possible to detect activity failures through the use of suitable failure detection functions.

In the ideal situation for an implementation of the proposed framework one can use an existing ECM-IS such as CCPro [Th15a] that is oriented at the compliance management processes described in Section 2. In addition to extending the system by the data structures described above obviously also the system needs to be extended by a corresponding active system service that is able to guide compliance officers and also to assist them for risk control purposes. At the present state the development of such an active system service is an ongoing task of a research project that targets a corresponding extension of the CCPro system.

5 Conclusions

The large amount of regulations and the threatening sanctions when a company does not comply with relevant laws imposes increasing pressure to companies to make sure that the compliance management obligations are carefully addressed. Especially the obligations are to be performed with awareness for activity failures that may cause non-compliance. A lot of the activity failures are related to the fact that environmental compliance management activities are performed by individuals and also groups of individuals. It is an unrealistic assumption to expect that through whatever measures a work practice can be achieved where human errors do not occur. However, for good reasons it can be assumed that through suitable computer-based guidance and assistance it is possible to minimize the number of failures in corporate work areas.

The monitoring framework described in this article is a first step towards an approach to extend environmental information systems by a new service that will guide compliance officers and also assist them in risk control tasks. A part from providing a complete view of the current compliance risk situation it is intended that the service is able to propose risk mitigation activities. It is expected that in many cases through these mitigation activities a future state of non-compliance can pro-actively be prevented.

Literature

[Ag06] Agrawal, R., Johnson, C., Kiernan, J., Leymann, F.: Taming Compliance with Sarbanes-Oxley Internal Controls Using Database Technology. *Proc. 22nd IEEE Int. Conf. on Data Engineering*, 2006, pp. 92-101.

[BM09] Butler, T.; McGovern, D.: A conceptual model and IS framework for the design and

adoption of environmental compliance management systems, Information System Frontiers, No. 14, Springer, 2009, pp. 221-235

[Bo09] Bonazzi, R. H.: Compliance Management is Becoming a Major Issue in IS Design, In A. S. D'atri, Information Systems: People, Organizations, Institutions, and Technologies, 2009, Springer, pp. 391-398

[FT09] Freundlieb, M., Teuteberg, F.: Towards a Reference Model of an Environmental Management Information System for Compliance Management. In *EnviroInfo 2009 (Berlin) Environmental Informatics and Industrial Environmental Protection: Concepts, Methods and Tools*. Berlin: Shaker Verlag 2009, pp. 129-138.

[Gu11] Gunningham, N.: Enforcing Environmental Regulation. *Journal of Environmental Law,* 23(2), 2011, pp. 169-201.

[HMB11] Haas de, H., Meerman, P., Bree de, M.: Compliance Management and System-Based Supervision, Proc. Ninth Int. Conf., Environmental Compliance and Enforcement, 2011, pp. 325-340

[Im95] IMA: Implementing Corporate Environmental Strategies, *Business Performance Management 67.* Institute of Management Accountants (IMA), Montvale, NJ, 1995.

[Im12] IMPEL: Compliance assurance through company compliance management systems 2011/04. European Union Network for the Implementation and Enforcement of Environmental Law (IMPEL), 2012.

[Ja72] Janis, I.: Victims of groupthink. Boston: Houghton Mifflin, 1972

[Ke03] Kerrigan, S. L.: A software infrastructure for regulatory information management and compliance assistance, PhD Dissertation. Stanford, CA, USA: Stanford University, 2003.

[Ma90] Martin, J., Information Engineering: Planning & Analysis, Book II. Englewood Cliffs, NJ: Prentice-Hall, 1990

[Me10] Melville, N.P.: Information Systems Innovation for Environmental Sustainability, MIS Quarterly, vol. 34, no. 1, 2010.

[SLM10] Salfner, F., Lenk., M., Malek, M.: A Survey of Online Failure Prediction Methods. *ACM Computing Surveys*, 42(3), March 2010, Article No. 10.

[TS09] Teuteberg, F. and Straßenburg, J.: State of the art and future research in Environmental Management Information Systems - a systematic literature review," in Information Technologies in Environmental Engineering, Proceedings of the 4th International ICSC Symposium Thessaloniki, Greece, Berlin, Springer, 2009, pp. 64-77.

[Th15a] Thimm, H.: IT-Supported Assurance of Environmental Law Compliance in Small and Medium Sized Enterprises. *Int. Journal of Computer and Information Technology,* 4(2), March 2015, pp. 297-305.

[Th15b] Thimm, H.: A Continuous Risk Estimation Approach for Corporate Environmental Compliance Management, Proc. IEEE 15th Int. Conf. on Environmental and Electrical Engineering, June, 2015, Rome, forthcomming

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Zwecke von betrieblichen Umweltinformationssystemen im Rahmen einer nachhaltigen Entwicklung

Andreas Moeller[1]

Abstract: In diesem Beitrag soll der Frage nachgegangen werden, welche Zwecke mit betrieblichen Umweltinformationssystemen (BUIS) verfolgt werden. Der Schwerpunkt liegt dabei bei der Verknüpfung zwischen Erkenntnisinteressen und Ansätzen der computergestützten Modellbildung und Simulation. Ausgangspunkt ist das weltweit bereits etablierte Life Cycle Assessment, das trotz des Erfolgs als alleinige methodische Grundlage von BUIS nicht ausreichen dürfte. Insbesondere werden wesentliche Erkenntnisse für die betriebliche Nachhaltigkeitsstrategien nicht angemessen abgedeckt. Eine Befassung mit den theoretischen Grundlagen des Life Cycle Assessments – insbesondere in Verbindung mit der Kostenrechnung – zeigt, dass eine sehr einfache Relation zwischen betrieblichen Entscheidungen und der konkreten Modellbildung hergestellt wird. Diese Beziehung führt zu einer wechselseitigen Verengung von Korridoren. Einerseits sind nur Effizienzerkenntnisziele abgedeckt, andererseits werden isolierte Ein-Produkt-Systeme analysiert. Damit können aber insbesondere Konsistenz und Gesamteffizienz vernetzter Wertschöpfungssysteme im Rahmen planetarer Grenzen nicht unterstützt werden. Im Ergebnis wird vorgeschlagen, dass sich die Modellbildungskonzepte von BUIS auf Entscheidungen und den Relationen zwischen ihnen beziehen sollte, die man daher in die Modellbildung einbeziehen sollte. Ein Vorbild könnten die Strategy Maps sein.

Keywords: Betriebliches Umweltinformationssystem, Life Cycle Assessment, Kostenrechnung, Identitätsprinzip, Strategy Maps

1 Einleitung

Stoffstromanalysen gelten als das Fundament nachhaltigkeitsbezogener betrieblicher Umweltinformationssysteme (BUIS) [HH95, Ra99]. Sie sind heute oft als Effizienzanalysen implementiert: Mit Hilfe von Life Cycle Assessments (LCA) [Gu02] werden die von Produkten und Dienstleistungen verursachten Umweltwirkungen abgeschätzt. Ausgangspunkt sind eine einzelne Entscheidung oder ein einzelner Entscheidungskomplex, etwa ein Produkt herzustellen oder eine Dienstleistung anzubieten. Über drei Stufen hinweg wird die Entscheidung über 1:1-Beziehungen mit Stoff- und Energieströmen verknüpft: (1) zwischen Entscheidung und Erkenntniszielen des Life Cycle Assessments, (2) zwischen Entscheidung und funktioneller Einheit (ein Nutzen) sowie (3) zwischen funktioneller Einheit und Referenzfluss[KG07]: ein Referenzfluss dient als Bezugsgröße für Effizienzanalysen.

[1] Leuphana Universität Lüneburg, Institut für Umweltkommunikation, 21335 Lüneburg, moeller@uni.leuphana.de

Diese einfachen 1:1-Beziehungen sind in der Praxis so klar, dass ihre Beschreibung eher als eine lästige Dokumentationspflicht empfunden wird: man trägt einen Standardtext im Feld „Goal & Scope“ ein. Das Vorgehen hat allerdings eine Nebenfolge: das Allokationsproblem bei Kuppelprozessen. Das Allokationsproblem hat die Eigenschaft, dass es wirklich gelöst werden muss. Ansonsten können die Berechnungsverfahren nicht eingesetzt werden: die „Technology Matrix“ ist nicht quadratisch [HS02]. In den Lehrbüchern zum Life Cycle Assessment nimmt das Allokationsproblem breiten Raum ein [KG07], und es verdeutlicht sich die Einsicht, dass sich die Lösung des Allokationsproblems nicht in den Naturwissenschaften findet, wenn sich ein Ansatz auf „physical relationships“ bezieht [EF01]. Mit dem Lösen des Allokationsproblems sollen vielmehr artifizielle Grenzen gezogen werden. Im Life Cycle Assessment nennt man die sich ergebenden Modelle Produktsysteme [KG07]. Bei diesen Modellen werden alle Stoff- und Energieströme isoliert, die für die Erbringung des Referenzflusses notwendig sind bzw. die von ihm verursacht worden sind.

Mit dem Allokationsproblem steht das Life Cycle Assessment nicht allein da. Es gibt eine lange Tradition in den Wirtschaftswissenschaften zum Umgang mit genau diesem Problem [Ri90, Kü92]. Die Erkenntnisse in der Kostenrechnung sollen dazu genutzt werden, neue Zugänge zu den Modellierungsgrundlagen von Stoffstromanalysen zu erschließen und aus ihnen ein Rahmenwerk für BUIS abzuleiten

2 Ansatzpunkte für Erkenntnisziele von BUIS

Ein radikaler Ansatz der Kostenrechnung besteht darin, die eingangs beschriebenen 1:1-Beziehungen in Frage zu stellen. Das hat bereits in den 1950er Jahren Riebel getan [Ri90, Ri92]. Er schlägt als Grundlage für Effizienzanalysen das Identitätsprinzip vor. Das Identitätsprinzip steht für ein Dreiecksverhältnis zwischen Entscheidungen sowie den ein-eindeutig zuzuordnenden positiven und negativen Wirkungen. Die Übertragung auf betriebliche Stoffstromanalysen soll in diesem Beitrag dazu dienen, die Möglichkeiten komplexerer Relationen zu untersuchen, weitere potentielle Zwecke von BUIS zu identifizieren und sie zu integrieren. Das Leitmotiv besteht darin, durch BUIS ein effizientes und konsistentes Wirtschaften innerhalb planetarer Grenzen zu unterstützen.

Wenn auch nun eine differenziertere Betrachtung möglich ist, bleibt der Ausgangspunkt doch unverändert: die betriebliche Entscheidung. Damit wird ein bestimmtes Selbstverständnis vorausgesetzt: Durch was oder durch wen werden die relevanten Veränderungen in unserer Gesellschaft angestoßen? Hier ist es die Annahme, dass Veränderungen auf Entscheidungen von Entscheidungsträgern zurückzuführen sind. Es wird von der Haltung ausgegangen, dass wesentliche Defizite im Bereich einer nachhaltigen Entwicklung auf Entscheidungen zurückzuführen sind, die Rationalitätsmängel aufweisen, welche ihrerseits auf fehlende Informationen und unzureichende Rationalitätssicherung zurückzuführen sind [We98].

2.1 Entscheidung und Information

Zunächst muss geklärt werden, in welchem Verhältnis Entscheidung und Informationsbedarf stehen, wobei hier unter einer Entscheidung ein unklare, wenig strukturierte Situation verstanden wird, die sich durch betriebliches Handeln in eine bestimmte Richtung weiterentwickelt. Die Kontexte haben unterschiedlichen Charakter:

(1) *Informationsversorgung für konkret anstehende betriebliche Entscheidungen im engeren Sinne:* Grundlagen der Datenverarbeitung sind dabei oft vergangenheitsbezogene Daten. Sie werden im Rahmen eines umwelt- und nachhaltigkeitsbezogenen Controllings verwendet. Abweichungen und damit Problemfelder werden identifiziert und abgeschätzt. Das Vorgehen kann als evolutionär charakterisiert werden. Gerade in Bezug auf die betriebliche Nachhaltigkeit und dazu anstehende Transformationen reicht der Blick in die Vergangenheit nicht aus. Die Datenverarbeitung hat daher zunehmend die Funktion der Modellierungsunterstützung. Mit anderen Worten: Die Datenverarbeitung wird durch Modellspezifikation und Modellparametrisierung ersetzt. Simulationsmodelle erlauben dann das Experimentieren mit neuen Strukturen und Prozesse.

(2) *Informationsversorgung für die betrieblichen Kommunikationsprozesse:* Der Kommunikationsbegriff bezieht sich dabei auf den Informationsaustausch zwischen verschiedenen Entscheidungsträgern (organisationsintern und -extern). Aus der Perspektive eines Unternehmens werden die Kommunikationspartner auch als Stakeholder bezeichnet. Die betrieblichen Kommunikationsprozesse sind aber nicht nur auf den Informationsaustausch beschränkt. Kommunikation ist auch die Grundlage organisationaler Verständigungs- und Lernprozesse. Wenn zum Beispiel von Pilotprojekten die Rede ist, dann erhofft man sich einen solchen Lernprozess.

(3) *Dokumentationsaufgaben:* Das BUIS unterstützt rechtlich vorgeschriebene und freiwillige Dokumentationspflichten. Gerade in Bezug auf die betriebliche Nachhaltigkeit verbindet sich mit der Dokumentation auch der Aufbau von Wissen. Solches Wissen ist eng verknüpft mit dem Herauskristallisieren betrieblicher Routine (Habitualisierung und Institutionalisierung).

(4) *Umweltschutz- und nachhaltigkeitsbezogene Geschäftsprozesse:* Hier geht es um die effektive und effiziente Unterstützung von neuartigen Routineprozessen. Das BUIS dient hier der Operationalisierung von nachhaltigkeitsbezogenen Strategien.

Die Ausführungen zeigen, dass der Begriff der Entscheidung hier sehr weit gefasst wird. Insbesondere sind nicht alle relevanten Entscheidungen darauf bezogen, direkt effizientes betriebliches Handeln einzuleiten. Oft haben die Entscheidungen vorbereitenden oder absichernden Charakter.

2.2 Information und Funktionen

Hier geht es darum, die Bezüge zwischen Erkenntniszielen und den positiven Wirkungen zu modellieren. Welchen Nutzen hat eine Organisation und / oder die Gesellschaft von einem Wertschöpfungsverbund? Das Life Cycle Assessment lässt an einzelne Produkte und Dienstleistungen denken, gleichwohl kann der Nutzen vielfältig sein.

Verschiedene Kategorien können anhand der verfolgten Nachhaltigkeitsstrategien einer Gesellschaft unterschieden werden: Effizienz, Konsistenz und Suffizienz [Hu94, Hu00, SP02]. Bezogen auf einzelne Unternehmen kann man mit Blick auf Suffizienz auf von den planetaren Grenzen sprechen [Ro09]. Darunter ist zu verstehen, dass mittelfristig dafür gesorgt werden muss, etwa durch eine entsprechende Gesetzgebung, dass global die planetaren Grenzen einzuhalten sind und dass sich daraus Herausforderungen für Unternehmen ergeben werden.

(1) *Produkte und Dienstleitungen:* Hier ist der Blick auf Effizienzfragen gerichtet [Sc08]: Produkte und Dienstleistungen sollen mit möglichst geringen Auswirkungen für die natürliche Umwelt hergestellt und genutzt werden. Einen besonderen Schwerpunkt bilden derzeit die Energie- und die Ressourceneffizienz. Hinzu kommt, dass die Kosten möglichst gering sein sollten. Ein Hersteller wird möglichst geringe Herstellungskosten anstreben, der Nutzer von Produkten und Dienstleistungen die „Total Costs of Ownership“ möglichst niedrig halten wollen.

(2) *Einhalten planetarer Grenzen:* In dieser Perspektive wird der Blick auf die Bereiche gerichtet, bei denen mutmaßlich planetare Grenzen überschritten werden [Ro09]. Die Effizienzstrategien mögen zwar zu einer Milderung führen, sind allein aber nicht hinreichend, wie die Diskussion um Rebound-Effekte immer wieder verdeutlicht [VV14]. Ausgangspunkt hier ist, dass die Einhaltung planetarer Grenzen auf kurz oder lang eine Nebenbedingung betrieblichen Handelns wird. Das BUIS hat in der Hinsicht eine vorbereitende und risikomindernde Funktion.

(3) *Konsistenz:* Konsistenz wird in der Nachhaltigkeitsdebatte auf den Stoff- und Energieaustausch mit der natürlichen Umwelt bezogen. Die absoluten Stoff- und Energieströme sollten im Ergebnis so ausfallen, dass die Ökosysteme nicht überlastet werden und ein eine Art „Open-Loop-Recycling“ mit Hilfe von Ökosystemdienstleistungen ermöglicht wird. Neben der absoluten Höhe des Stoff- und Energieaustausches kommen auch Fragen der Abbaubarkeit, Toxizität usw. hinzu, Fragestellungen also des betrieblichen Umweltschutzes seit vielen Jahrzehnten. Eine Gestaltungsvariable bei der Herstellung von Konsistenz ist die Kreislaufführung von Stoffen innerhalb des anthropogenen Systems. In der Perspektive der Konsistenz kann man entsprechend auch den Stoff- und Energieaustausch zwischen verschiedenen Produktsystemen untersuchen. Gefordert wird die Integration der Produktsysteme: Aus isolierten Produktsystemen wird ein integrierter Produktionsverbund. Produkte und Dienstleistungen sind also so zu gestalten, dass eine möglichst nahtlose Integration in den gesamtgesellschaftlichen Wertschöpfungsverbund möglich wird. Dabei kann man sich auch schlüssiger Entwurfsleitbilder wie Cradle-to-Cradle [MB02] bedienen. Das

Ziel eines BUIS besteht darin, Informationen dazu zur Verfügung zu stellen, ob und wie eine konsistente Integration möglich ist.

Offensichtlich ist, dass die Aufgaben von BUIS vielfältig sind und ganz unterschiedliche Betrachtungsgegenstände einer Analyse und Einschätzung bedürfen. Man könnte für alle anstehenden Fragestellungen verschiedene, nicht miteinander Daten austauschende „Apps“ entwickeln. Der Anspruch an ein Informationssystem geht allerdings weiter. Zwar sind verschiedene Auswertungsinstrumente denkbar – wenn man so will: Apps. Aber sie sollten auf zwei Ebenen integriert werden: erstens eine gemeinsame Datenbasis und zweitens auf der Ebene der Entscheidungen und den Beziehungen zwischen ihnen. Die Vorüberlegungen sollen im Folgenden dazu genutzt werden, Gestaltungsleitbilder für BUIS zu entwerfen. Als Zentralbegriffe sollen die Stoffstrommodelle und die Strategy Map dienen.

3 BUIS als Rahmenwerke für nachhaltigkeitsbezogene Strategy Maps

Es haben sich eine ganze Reihe von Ansätzen, Prinzipien und Regeln verfestigt, die eine betriebliche nachhaltige Entwicklung einleiten und fördern sollen. Diese sind von einigen Wissenschafts-Communities ausgearbeitet und standardisiert worden. Die Aufgabe der Umweltinformatik sollte dann im Sinne einer „instrumentellen“ Wissenschaft sein, aus den Vorgaben die notwendige IT abzuleiten. Dies ist dann gut gelungen, wenn man ein passendes „Community-Tool“ entwickelt hat. Der Ansatz ist im Fall des Life Cycle Assessments sehr erfolgreich gewesen.

Aus den Ausführungen des Abschnitts 2 folgt allerdings, dass sich ein einziger methodischer Ansatz allein als zu einseitig erweisen dürfte. Beim Life Cycle Assessment beispielsweise richtet sich, wie die Ausführungen bereits verdeutlicht haben, der Blick auf die Effizienz einzelner, isolierter Produkte oder Dienstleistungen. Verbundwirkungen oder auch die absoluten planetaren Grenzen werden mit dem Ansatz nicht erfasst. Im Gegenteil kann nicht ausgeschlossen werden, dass dann, wenn von Win/Win-Situationen gesprochen wird, das zentrale Problem der Effizienzstrategie adressiert wird: Reboundeffekte [VV14]. Man könnte von einer Dialektik der Effizienzstrategie sprechen: Die guten Absichten, mit denen man begonnen hat, verkehren sich an Ende ins Gegenteil; Veränderung im Sinne einer nachhaltigen Entwicklung wird gerade verhindert. Der alleinige Fokus auf Effizienz könnte sich damit gegen betriebliche Strategieelemente richten, die auf die langfristige Existenzsicherung gerichtet sind.

Hier wird der Vorschlag unterbreitet, in dem Zusammenhang den Grundgedanken der Strategy Maps [Kn04] für BUIS fruchtbar zu machen und mit Stoffstrommodellen als Datengrundlage zu verbinden (stoffstrombasierte BUIS [Mö00]). Verschiedene Auswertungen von BUIS werden über Strategy Maps miteinander in Beziehung gesetzt. Zugleich kann anhand der Strategy Maps die Konsistenz der Informationsverarbeitung in Bezug auf die betrieblichen Strategien untersucht bzw. nachgewiesen werden.

Die Strategy Maps haben für das stoffstrombasierte BUIS die Funktion einer Klammer: alle Komponenten eines BUIS haben den Zweck, die Strategy Maps mit Daten zu füllen. Mit dem Vorschlag verbunden ist demnach eine 3-Schichten-Architektur des BUIS mit (1) der obersten Ebene der Strategy Maps, (2) der Ebene der Auswertungen für einzelne Entscheidungen sowie (3) der untersten Ebene der Stoffstrommodelle. Im Folgenden sollen die Ebenen vorgestellt werden:

3.1 Stoffstrommodelle als Grundrechnung

Die 3-Schichten-Architektur stellt besondere Anforderungen an die Modellierung und Datenverarbeitung: Wenn verschiedene Modellierungen mit Hilfe eines übergeordneten Konzepts in einen Zusammenhang gestellt werden sowie Aggregationen und Vergleiche vorgenommen werden, dann müssen die Modelle auf der untersten Ebene konsistent erstellt und ausgewertet werden. Mit anderen Worten: BUIS müssen einen softwaretechnischen Rahmen dafür bieten, dass eine Integration der Daten möglich wird. Das Thema ist nicht neu in der Umwelt- bzw. Wirtschaftsinformatik. Im Falle der IT für betriebliche Nachhaltigkeit gibt es allerdings eine einzigartige Bezugsebene, die der wirklichen materiellen Stoff- und Energieströme [FK98, FH99]. Analog zur finanziellen Dimension bei der Balanced Scorecard [KN01] bildet sie die „Ergebnisdimension“ betrieblicher Nachhaltigkeitsaktivitäten ab.

Die Forschung auf dem Gebiet der stoffstrombasierten BUIS ist darauf ausgerichtet, die Modellierungsansätze so universell auszulegen und mit verschiedenen Auswertungsmethoden zu kombinieren, dass das gesamte Informationssystem auf einem (konsistenten) Stoffstrommodell basieren kann [Mö00]. Das Stoffstrommodell wäre dann etwas, was Riebel im Rahmen des von ihm entwickelten Kostenrechnungssystems als eine Grundrechnung bezeichnet [Ri90, Kü94]. Eine solche Grundrechnung zeichnet sich dadurch aus, dass eine 1:n-Beziehung zwischen Grundrechnung und Auswertungsrechnungen etabliert wird.

3.2 Einzelentscheidungen und Auswertungsrechnung

In Bezug auf das Konzept der Stoffstromnetzte [Mö00] hat sich bereits ein ganzer Katalog an Auswertungsrechnungen etabliert: Life Cycle Assessment als Auswertungsrechnung, eine verbundbezogene Kostenrechnung, Material Flow Cost Accounting (MFCA), Mehrperiodenmodelle zu Bestandsentwicklungen („anthropogene Lager“) usw. Auch werden Marginalanalysen unterstützt, die in größeren Modellen auch die Fragen zur Reichweite von Veränderungen erlauben (Verbundwirkungen von Entscheidungen). Anhand der Nachhaltigkeitsstrategien können 2 verschiedene Arten von Auswertungsrechnungen unterschieden werden:

(1) Auswertungsrechnungen, die Auskunft geben sollen zu absoluten Stoff- und Energieströmen, ihrer Zusammensetzung, ihrer „Anschlussfähigkeit“ in der natürlichen

Umwelt oder anderen Wertschöpfungssystemen, ihrem Verbleib in der natürlichen Umwelt, ihrer Dynamik und mit ihnen verbundenen Bestandsentwicklungen.

(2) Auswertungsrechnungen, die sich auf Effizienzfragen beziehen. Hier bildet das Life Cycle Assessment einen geeigneten Rahmen, der auch Kostenrechnungen in verschiedenen Varianten mit einschließt. Modellierung und Probleme der Modellierung sind bezogen auf anthropogene Stoff- und Energieströme weitgehend identisch (einschließlich Zurechnungsprobleme bei Kuppelproduktion, Umgang mit fixen Anteilen usw.), allerdings sind die Bilanzgrenzen unterschiedlich. Da kann man unterscheiden (1) den gesamten Produktlebenszyklus für die Abschätzung von Umweltwirkungen (LCA), (2) die betrieblichen Grenzen für die Kostenrechnung in der Produktion sowie (3) die Nutzungsphase bei der Abschätzung der gesamten Kosten der Nutzung (Total Cost of Ownership). Solche Effizienzrechnungen können als Auswertungsrechnungen von periodenbezogenen Stoffstrommodellen implementiert werden [Mö00].

Mit den Effizienzanalysen sind besondere Herausforderungen verbunden, denn sie gehen von isolierten Produktsystemen aus. Auf einen Ausweg deutet das Identitätsprinzip Riebels hin. Gleichwohl führen die Überlegungen von Riebel und ihre Übertragung auf BUIS zu einer „Atomisierung“ der Informationseinheiten: Jede betriebliche Entscheidung wird hinsichtlich ihrer direkten positiven und negativen Wirkungen untersucht.

Beim Consequential LCA [EW04, Fr06, BP07] lässt man sich von analogen Grundgedanken leiten, wobei die Überlegungen aus der Kostenrechnung derzeit kaum Beachtung finden. In einem Punkt allerdings geht das Consequential LCA über die rein betriebliche Perspektive der Kostenrechnung hinaus. Sie will auch die „market-mediated consequences“ erfassen. Das ergibt sich daraus, dass beim Life Cycle Assessment oft gar keine einzelbetriebliche Perspektive eingenommen wird. Fragestellungen des Life Cycle Assessments sind oft mit generischen Produkten oder Produktgruppen verbunden. Verändern sich die Lebenszyklen für ganze Produktgruppen, dann hat dies nicht mehr nur Wirkungen auf der materiellen Ebene der jeweiligen Produktsysteme. Einbezogen werden daher auch Modellierungsansätze aus den Wirtschaftswissenschaften etwa zu Angebot und Nachfrage. Die Ansätze des Consequential LCA zu „market-mediated consequences“ führen zu Hybridmodellen. Die Hybridmodelle zeichnen sich dadurch aus, dass nicht alle Konsequenzen allein auf der materiellen Ebene abgebildet werden können.

3.3 Strategy Maps als Integrationsrahmen

Von den Modellierungsansätzen für die „market-mediated consequences“, die eine Erweiterung der Grundrechnung darstellen, ist die bereits erwähnte Bündelung der atomisierten Informationseinheiten und Entscheidungen zu größeren Einheiten abzugrenzen: Oberhalb der einzelnen Entscheidungen wird abgebildet, wie die verschiedenen, ineinandergreifenden Entscheidungsprozesse zum unternehmerischen Erfolg beitragen. Hier kommen die Überlegungen zu den Strategy Maps ins Spiel [Kn04]. Strategy Maps dienen dazu, verschiedene betriebliche Entscheidungen und Aktivitäten

systematisch auf den Unternehmenserfolg auszurichten. Diese Entscheidungen und Aktivitäten stehen in einem unterschiedlichen Verhältnis zum betrieblichen Erfolg: sie können direkter oder indirekter Natur sein. Die Herausforderung besteht darin, die indirekten Beiträge abbilden und untersuchen zu können. So ist auch die Strategy Map definiert: Sie beschreibt, wie ein Unternehmen Werte erzeugt und in welchen Ursache-Wirkungszusammenhang verschiedene unternehmerische Entscheidungen und Aktivitäten stehen, wobei die „finale“ Wirkung die Steigerung des langfristigen „Shareholder Values“ sein soll [KM04].

Dabei können – in der Regel vier – Perspektiven unterschieden werden [KN02, KN04]: Die Finanzperspektive, die Kundenperspektive, die der internen Prozesse sowie schließlich die Perspektive des Lernens und Wachsens [KN04]. Schaltegger und Dyllick schlagen in Bezug auf eine nachhaltige Entwicklung als weitere Perspektive die Gesellschaftsperspektive vor, welche die Bezüge zu gesellschaftlichen Anspruchsgruppen eines Unternehmens abdeckt, die nicht Kunden oder Shareholder sind [SD02].

Die grundsätzlichen, bei den Strategy Maps unterstellten Ursache-Wirkungsbeziehungen nehmen ihren Anfang bei der Lern- und Wachstumsperspektive. Diese wirkt ein auf die Ebene der internen Prozesse, die wiederum auf die der Kundenbeziehungen und diese schließlich auf die der finanziellen Wirkungen. Es ergeben sich Prozessnetze von zusammenhängenden Entscheidungen und Folgeentscheidungen [Ri92].

In solche Ursache-Wirkungsgefüge lassen sich auch nachhaltigkeitsbezogene betriebliche Entscheidungen und Aktivitäten einordnen. So können Pilotprojekte dazu dienen, das organisationale Lernen anzuregen, oder der Ersatz nicht-nachhaltiger Produktionsprozesse durch nachhaltigere könnte den Zweck haben, zukünftige gesellschaftliche Anforderungen aufzugreifen und aufzufangen (z.B. Einhalten planetarer Grenzen oder konsistente Produktionsprozesse im Sinne der Recyclingfähigkeit von Produkten usw.).

Die Atomisierung der Entscheidungen könnte dabei eine Zwischenstufe der Aggregation erforderlich machen, denn die Strategy Maps und schließlich die Balanced Scorecard werden nur von wenigen hochaggregierten Kennzahlen gespeist. Eine solche Aggregation hat Riebel bereits mit den Bezugsobjekthierarchien vorgeschlagen [Ri92], die sich keineswegs auf die Aggregation von Produkten und Dienstleistungen bezieht. Riebel erwähnt auch Kunden, Aufträge u.ä. Derart universell ausgelegte Bezugsobjekthierarchien lassen auch „Objekte“ der betrieblichen Nachhaltigkeit zu.

4 Zusammenfassung

Der Vorschlag dieses Beitrag ist, die Grundgedanken zu den Grundrechnungen und zu den Strategy Maps aufzugreifen, um aus ihnen einen doppelten Integrationsrahmen für betriebliche Umweltinformationssysteme abzuleiten. Doppelt deswegen, weil (1) die verschiedenen atomisierten Informationseinheiten zu Gesamtbildern integriert werden und (2) von einer gemeinsamen Grundrechnung ausgegangen werden kann; diese bezieht

sich auf die absoluten Stoff- und Energieströme sowie die Entwicklung der Bestände in einem Wertschöpfungsverbund.

Ein solcher doppelter Integrationsrahmen erfordert Auswertungsinstrumente von Stoffstrommodellen, die sich konsequenter an den einzelnen Entscheidungen orientieren. Derzeit dienen Stoffstrommodelle in erster Linie dazu, Daten für das Life Cycle Assessment und eine Vollkostenrechnung zur Verfügung zu stellen. Das ändert sich mit dem doppelten Integrationsrahmen: Consequential LCA in einem verallgemeinerten Sinne und Einzelkostenrechnungen können als Zwischenstufen zum Einsatz kommen. Entscheidungsgegenstände können dann neben Produkten und Dienstleistungen sein: Produktionsprozesse, Investitionsentscheidungen, Fortbildungsmaßnahmen usw.. Für strategische Ziele wie die absolute Deckelung oder die absolute Reduktion von Umweltwirkungen kann die Rationalität gesichert werden.

Literaturverzeichnis

[BP07] Bauer, C. Poganietz, W.-R.: Prospektive Lebenszyklusanalyse oder die Zukunft in der Ökobilanz, Technikfolgenabschätzung – Theorie und Praxis 16 (3), 17-23, 2007.

[FK98] Fischer-Kowalski, M.: Society's Metabolism – The Intellectual History of Material Flow Analysis, Part I, 1860 – 1970, Journal of Industrial Ecology 2 (1), 61 – 78, 1998.

[FH99] Fischer-Kowalski, M., & Hüttler, W.: Society's Metabolism – The Intellectual History of Material Flow Analysis, Part II, 1970 – 1998, Journal of Industrial Ecology 2 (4), 61-78, 1999.

[Fr06] Frischknecht, R.: Modelling of product systems in life cycle inventory analysis: Synopsis of Attributional and Consequential LCI Models – Properties and Differences, Bericht an das Netzwerks Lebenszyklusdaten, 2006.

[EF01] Ekvall, T., Finnveden, G.: Allocation in ISO 14041—a critical review, Journal for Cleaner Production 9, 197–208, 2001.

[EW04] Ekvall, T. Weidema, B.: System Boundaries and Input Data in Consequential Life Cycle Inventory Analysis, International Journal for Life Cycle Assessment 9 (3), 161-171, 2004.

[Gu02] Guinée, J.B. (Final Editor): Handbook on Life Cycle Assessment – Operational Guide to the ISO Standards, Dordrecht, Boston, London, 2002.

[HH95] Haasis, H.-D., Hilty, L.M., Kürzl, H., Rautenstrauch, C.: Anforderungen an Betriebliche Umweltinformationssysteme (BUIS) und Ansätze zur Realisierung, in: Haasis, H.-D., Hilty, L.M., Kürzl, H., Rautenstrauch, C. (Hrsg.): Betriebliche Umweltinformationssysteme (BUIS) – Projekte und Perspektiven, Marburg, 1995.

[HS02] Heijungs, R, & Suh, S.: The computational structure of life cycle assessment, Dordrecht, 2002.

[Hu94] Huber, J.: Nachhaltige Entwicklung durch Suffizienz, Effizienz und Konsistenz, in: Fritz, P., Huber, J., Levi, H.W. (Hrsg.): Nachhaltigkeit in naturwissenschaftlicher und

sozialwissenschaftlicher Perspektive, Stuttgart, 31-46, 1994.

[Hu00] Huber, J.: Industrielle Ökologie. Konsistenz, Effizienz und Suffizienz in zyklusanalytischer Betrachtung, in: Kreibich, R., Simonis, U. E.(Hrsg.): Global Change, Berlin, 109–126, 2000.

[KG07] Klöpffer, W., Grahl, B.: Ökobilanz (LCA) – Ein Leitfaden für Ausbildung und Beruf, Weinheim, 2007.

[KN01] Kaplan, R.S., Norton, D.P.: Die strategiefokussierte Organisation – Führen mit der Balances Scorecard, Stuttgart, 2001.

[KN04] Kaplan, R.S., Norton, D.P.: Strategy Maps – Converting Intangible Assets into Tangible Outcomes, Boston, 2004.

[Kü92] Küpper, H.-U.: Theoretische Grundlagen der Kostenrechnung, in: Männel, W. (Hrsg.): Handbuch Kostenrechnung, Wiesbaden, 1992.

[Kü94] Küpper, H.-U.: Paul Riebel – Wichtige Perspektiven für die Kostenrechnung, krp Kostenrechnungspraxis 1/94, 40-42, 1994.

[MB02] McDonough, W., Braungart, M.: Cradle to Cradle: Remaking the Way We Make Things, New York, 2002.

[Mö00] Möller, A.: Grundlagen stoffstrombasierter Betrieblicher Umweltinformationssysteme, Bochum, 2000.

[Ra99] Rautenstrauch, C.: Betriebliche Umweltinformationssysteme – Grundlagen, Konzepte und Systeme, Berlin, Heidelberg, New York, 1999.

[Ri90] Riebel, P.: Einzelkosten- und Deckungsbeitragsrechnung, 6. Auflage, Wiesbaden, 1990.

[Ri92] Riebel, P.: Einzelerlös-, Einzelkosten- und Deckungsbeitragsrechnung als Kern einer ganzheitlichen Führungsrechnung, in: Männel, W. (Hrsg.): Handbuch Kostenrechnung, Wiesbaden, 1992.

[Ro09] Rockström,, J. et al.: A safe operating space for humanity, Nature 461, 472 - 475, 2009.

[Sc08] Schmidt, M.: Die Bedeutung der Effizienz für Nachhaltigkeit – Chancen und Grenzen, in: Hartard, S., Schaffer, A., Giegrich, J. (Hrsg.): Ressourceneffizienz im Kontext der Nachhaltigkeitsdebatte, Baden-Baden, 31-46, 2008.

[SD02] Schaltegger, S., Dyllick, T. (Hrsg.): Nachhaltig managen mit der Balanced Scorecard, Wiesbaden, 2002.

[SP02] Schaltegger, S., Petersen, H.: Strategisches Umweltmanagement und Öko-Marketing, Centre for Sustainability Management (CSM), Lüneburg, 2002.

[VV14] Vivanco, D.F., van der Voet, E.: The rebound effect through industrial ecology's eyes: a review of LCA-based studies, International Journal for Life Cycle Assessment 19, 1933–1947, 2014.

[We98] Weber, J.: Einführung in das Controlling. Stuttgart, 1998.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Integrating IT-for-Green Project through Web Services with the System of Sustainability Performance Evaluation

Jorge Marx Gómez[1], Frank Medel-González[2]

Abstract: Corporate sustainability and sustainability management have become a key issues in modern organizations to achieve a balanced and sustainable development. Nowadays IT can support the sustainability behaviors in business processes to improve the sustainability performance in organizations. IT-for-Green is a project that proposes a new generation of Corporate Environmental Management Information Systems allowing to incorporate the strategic sustainability integration. The main motivation of this to provide a first approach of how to achieve communication, interoperability and integration with other platforms to improve the cover areas of IT-for-Green and accomplish one of the goal of this project "the integration with other applications". The paper explores the communication and integration of IT-for-Green with the System of Sustainability Performance Evaluation using web services as a platform. A first web services definition was designed.

Keywords: sustainability management, solutions integration, web services

1 Introduction

The recent growth in corporate sustainability made organizations consider this area as a key success factor that must be managed [Dy02] and [Sc11]. IT plays an important role in sustainability management, specifically in sustainability performance evaluation [MMW13], [Me10] and [PP09]. Although IT has environmental impacts during its life-cycle, as a positive part, IT supports eco-controlling and efficiency in organizations.

The last fifteen years have raised the support of IT to the environmental and sustainability behavior in business organizations [Me10], [RB13] and [MA12]. Different concepts have been popularized through the academic and business world, e.g. Green Information Systems (IS), Green IS & IT, Green computing, Green IT and IT-for-Green. All are related to first-order effects (negative environmental impact of IT) and second-order effects (positive impact of IT in business processes). IT-for-Green is one of the newest concepts of the second-order effects and refers to the positive impact of using IT on business and economic processes. This perspective considers IT as part of the solutions to eco-sustainability [MA12].

Established tools for strengthening IT support are Corporate Environmental

[1] Department of Informatics, Very Large Business Applications (VLBA), University of Oldenburg, 26129 Oldenburg, Germany, Jorge.Marx.Gomez@uni-oldenburg.de
[2] Universidad Central "Marta Abreu" de Las Villas, Faculty of Industrial Engineering, Santa Clara,Villa Clara, Cuba, frankmedel@uclv.edu.cu

Management Information Systems (CEMIS), but those are not sufficient yet to achieve the strategic sustainability integration [He07] and [Wo08]. For that reason, the project IT-for-Green started in 2011, IT-for-Green covers the complete product life cycle from input to transformation to output. It proposes a new generation of more strategic CEMIS, which should be able to support the company's decision makers in all stages of product life cycle [It13]. Organizations need to track their sustainability goals and the goals of all their branches in a continuous way. The sustainability indicators are a good tool to compare sustainability business performance in different branches by setting an internal sustainability benchmarking. Managers prefer the condensed information for a quick understanding of the whole business picture, identifying setbacks and progress related to the overall performance.

The main motivation of this research was oriented to fill an important gap in relation with one goal of IT-for-Green project "the integration of IT-for-Green with other applications". The business need to respond quickly to changes in the business environment in which they operate. Isolated applications tend to become problematic after a few years in use, then their integration in changing IT-environment becomes more difficult. The proliferation of web services within the last two years enables organizations to assimilate software and services from different companies and locations into an integrated service capable of streamlining important processes [LG15].

The goal is to provide a first approach of how to achieve communication, interoperability and integration of those platforms; to add further functionalities from other applications using web services as a platform to improve the cover areas of IT-for-Green CEMIS. Also further contribution to the internal benchmarking of organizations through the strengthening and upgrading of reporting functionalities of other applications.

Communication and reporting are key elements of an organization's sustainability management. For that reason the idea of System of Sustainability Performance Evaluation (SySPE) developed for Cuban organizations [Me13], can be included as a future extension of web services collection for the IT-for-Green project.

SySPE is a tool to support the storage, retrieval management and integration of different sustainability indicators. The organizations' managers set the indicators' goals, related to the business performance into the application. SySPE helps to calculate the Corporate Index of Sustainability Performance (CISP) (see [Me13]). This index serves managers to discover which is the overall compliant of sustainability business goals and includes the perception of different stakeholders. The application allows the graphical representation of CISP and visualizes the improvement of the potentials of indicators to redirect the business managers' efforts.

All these functionalities can help to track business sustainability behavior and continuously improve the internal benchmarking. The main objective of this paper is to integrate the IT-for-Green project with SySPE, specifically to the module, which is related to sustainability reporting and dialogue (Module 3). The paper explores the

possibility of effectively harmonize these two systems, using web service with Zend Framework to add some future SySPE functionalities. The validation of the web service was made through Web Services Description Language (WSDL) and SOAP messages validation.

The paper outline is structured in (i) introduction, (ii) theoretical review of Sustainability management and IT support, (iii) IT-for-Green and SySPE descriptions, (iv) web services definition and practical example and finally (v) future steps and conclusion.

2 IT Support for corporate environmental management and sustainability

The business has a great responsibility in the process of transition to a more sustainable development (SD). SD is a social concept, and is being increasingly applied as a business concept under the name of corporate sustainability [St05].

In recent years, some tools had been developed to help organizations in the long path of sustainability. The role of IT to support environmental and sustainability management and reporting had being strengthened for the IT capabilities [Me10], [RB13], [MA12], [Ca13], [HB14] and [Hi14]. Decision Support Systems are emerging as a suitable solution in the field of sustainability planning and control of complex systems [Fi08]. According to [RB13] specialized tools are: SAP Sustainability Performance Management (SuPM), Enablon SD-CSR, SoFi and credit360, and STORM. Al these solutions are on the market as a commercial solutions difficult to access for countries like Cuba.

Another important tool is the project “Solution and Services Engineering for Measuring, Monitoring, and Management of Organizations’ Environmental Performance Indicators” (OEPI). This is an international research project supported and funded by the European Commission within the Seventh Framework Program [Ja12] which is related to environmental performance indicators. A fundamental goal of the OEPI is to bridge the gap between various sources and types of environmental information and users of different backgrounds by providing an integrated information source [Ja12]. OEPI provides to business users with an inter-organizational platform and tools to: provision and share environmental performance indicators across the chain and incorporate them in intra- and inter-organizational processes.

Other applications of IT solutions to support the relation between the organizations and the environment are : 1) ProPlaNET: a web based tool which supports sustainable project planning based on e-Participation and Web 2.0 [Gi10]; This web provides a comprehensive framework to deal with the decisions support in relation to the planning process including a large quantity of indicators, stakeholders interests and the transparency of the decision process 2) SIMASE (acronym in Portuguese): Information System for Environmental and Corporate Sustainability Monitoring, is a software

framework for social and environmental monitoring for diagnosis of organizations in terms of sustainability [TP13].

All of the above, evidences how the role of information technologies has increased in recent years to support environmental and sustainability information in order to support business decisions and accomplish organizations´ sustainability goals.

According to [It13] conventional CEMIS are not sufficient to achieve sustainability integration, whereas a new generation of more strategic CEMIS will be able to support the company's decision makers. IT-for-Green enrich traditional CEMIS using an integrated approach of handling processes [So13] using workflows.

3 Methodology

Nowadays applications tend to become problematic after a few years in-use and being integrated in a changing IT-environment. The business need to respond quickly to changes in the business environment in which they operate.

IT-for-Green with their modular approach allows add SySPE functionalities to improve the cover areas of IT-for-Green CEMIS and to take advantages of their functionalities.

The research problem was identified as: the necessity to make interoperable IT-for-Green with other applications to cooperate and use in relation with distributed applications in an open environment.

The investigation was performed following three main steps through exploration and integration of available technologies to produce the artifact. The first step is oriented to review the characteristics of the applications, functionalities and requirements. The second step consists in web service building; in this step the web service architecture is defined based in the previous overview of the application and a first example of web service was built. The third step is to validate the web service through WSDL and SOAP messages validation.

The firsts step was accomplished with the description of the two systems. Later a web service definition and design and finally the validation of the web service.

4 IT-for-Green project as a CEMIS solution

The primary goal of this project is "increasing the environmental friendliness of companies and their processes by means of ICT" [It13]. IT-for-Green proposed to research and create a new generation of Corporate Environmental Management Information Systems (CEMIS) which is able to support the company's decision makers.

The system is built in a modular manner [Ra11] and [Ma13] the modules are:

Green-IT: Organizations have to deal with energy efficiency as a relevant element to reduce their IT infrastructure carbon footprint and the potential of their climate change impact, through the optimization of their electricity grid. This module is oriented to support energy efficiency and data modelling; it helps to calculate the energy requirements of a data center and compares the results with reference data to optimize energy use and costs.

Green Production and logistics: This module gives insight on two basic CO_2 producing systems, namely production and logistics. For both systems there is a subsystem that models the existing processes and non-existing processes, so both can be compared to each other and to those of other companies.

Sustainability reporting and communication: The goal of this module is to collect and manage information about the real contribution of companies to sustainable development and stakeholders' interaction. The module handles economic, social and environmental information, necessary to current and future stakeholder demands. Reports are elaborated with the accepted guidelines GRI G3, but also other kind of reports can be transformed into a schema to be generated by the application.

IT-for-Green Next Generation CEMIS is built in a modular way follows the Service-Oriented Architecture (SOA). The modularization of IT-for-Green serves to different purposes:

1. Better integration in new IT environments, with IT-for-Green's modules only a module has to be renewed, not the whole application.
2. The possibility of modules customization in relation with the enterprise needs (they don't need to acquire the whole package);
3. The easy development, a modular system with loose coupling is easier to develop and maintain.

The platform is designed to be open and extensible for new nodules and services through a workflow-based and service-oriented platform [Ra11] and [Ma13]. According to [So13] the runtime environment comprises of the core building blocks or components: a Green Service Mall, a Workflow Engine, an Event Engine and a user interface, to handle the internal and external services and business processes.

Green Service Mall is the component that deals with web services, specifically it is a service repository where the web services can be published and discovered by consumers to satisfy their necessities.

The Workflow Engine is oriented to map the business process and it allows different tasks such as adding, updating, and deleting workflows; it is responsible for the system workflow execution and management, interacting with the service consumer and the workflow editor.

The Event Engine is a component of the CEMIS with the main task to compare pre-

established (environmental) requirements (e.g. water consumption, CO2 emissions, etc.) with the current variables' performance and to detect possible violations and generate warnings and alarm messages automatically. The Event Engine is composed by different subcomponents like: management, executor, timer, condition model and Data Access Objects (DAO).

The User Interface is the component layer between the CEMIS and the users and can be operated using any kind of web browsers.

One of the goals of IT-for-Green project is oriented to "integrate the prototypes into the surrounding web service infrastructure and the import of sustainability indicators and the generation of schemes [So13]. Actually the integration of IT-for-Green with other applications is an uncover objective.

5 System of Sustainability Performance Evaluation

SySPE is an informatics solution born in 2012, impelled by the necessity of the Cuban energy sector -- as representation of Cuban business organizations -- to respond to internal and external pressures derived from the inclusion of Sustainable Development concept at business levels, to support decisions, associated with sustainability performance and provide an internal benchmark and report tool to satisfy stakeholders information requirements. SySPE supports social, economic and environmental indicators related directly to organization performance distributed over key areas. Taking this into account, the idea of the Sustainability Balanced Scorecard (SBSC) was used, to pursuit the balance among the perspectives and the economic, environmental and social pillars.

SySPE has three modules. Data collection is related to the collection and storage of indicators information defined by business managers and regulatory standards. The BSC perspectives definitions belong to this module. Other actions are the update and elimination of information; those actions will be restricted to a small group of users that could interact with the module system.

Indicators aggregation: this module allows setting the sustainability indicators defined over the SBSC perspectives and assign weights for indicators and perspectives to calculate the Corporate Index of Sustainability Performance (CISP) see [Me13]. The CISP idea is synthesized in an index of the progress or setbacks in corporate sustainability performance, to verify simple and continuously if the managerial efforts, organizational management instruments and environmental training are translated into a better or worse sustainability performance. The third module for Graphic representation and report generation, allows users and stakeholders to visualize the behavior of CISP and sustainability indicators in a period and represent graphically the behavior of indicators and indexes.

The technologies used for application development were the MySQL, Propel as Object

Relational Mapping, Zend Framework (ZF) was used as rapid web development framework of PHP 5 implementing the Model-View-Controller (MVC) design pattern, which is the best current practice to develop modern web applications [Gi10]. The graphic representation of reports was implemented using the Business Intelligence and Reporting Tools (BIRT).

SySPE solution intents to cover a poor explored area in Cuban organizations; linked to the support of IT to sustainability performance management and business sustainability benchmark. The main goal is to support sustainability data and indicators to guide business managers and stakeholders to redirect sustainability issues efforts.

6 Web Services for Solutions Integration

SySPE and IT-for-Green are two solutions that support organizations in relation with the environment and sustainability. Since the design point of view they run in different technologies and pattern designs. IT-for-Green uses a Service Oriented Architecture design pattern and SySPE implements a Model-View-Controller pattern. In order to upgrade the IT-for-Green solution sustainability reporting and dialogue module, it is proposed to take advantage of SySPE functionalities and reporting capabilities. To achieve this goal one question should be answered: How to integrate two systems during the runtime effectively? A method of data exchange is needed.

An effective solution is a web service as a way to expose the SySPE functionalities and make it available through standard web technologies to facilitate application integration reducing applications heterogeneity. Web services promote the specification-based cooperation and collaboration among distributed applications in an open environment [Xi05]. Web services have become a widely used form of adding depth to online applications and allow developers write applications that are interoperable with external services located anywhere in the world [Pa09]. Web services use different web standards like XML and SOAP to tag and transfer the data, the WSDL is used for describing the services available and the Universal Description, Discovery and Integration (UDDI) is used for list what services are available [Al04].

IT-for-Green have a web service provider, which allows handle with the new services which are not available in the stock version of the CEMIS [Ma13] and a special controller can be implemented using Zend Framework to integrate through web services with SySPE To establish the integration between those applications, a basic web service architecture was defined (Fig. 1).

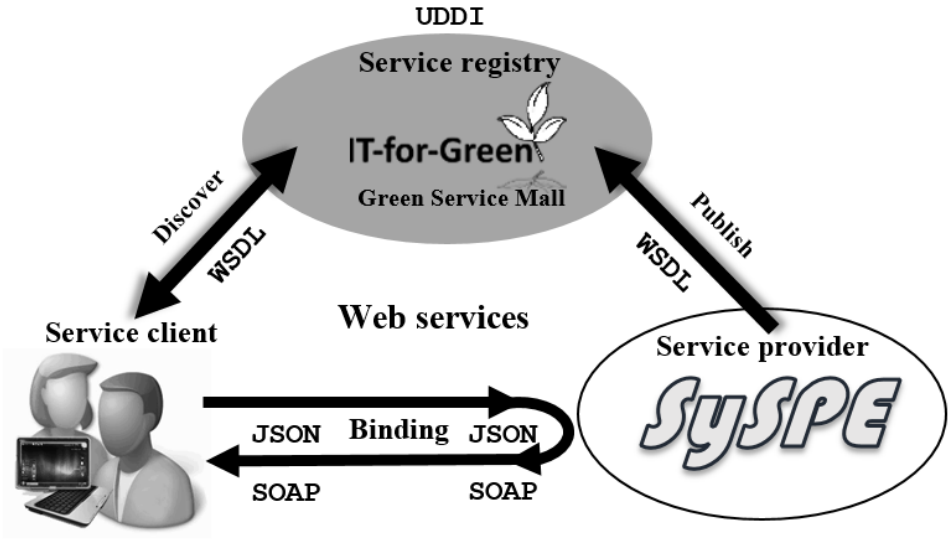


Fig. 1: Web services architecture.

It shows how the service client contact and discover the web service in service registry (Green service mall). The service registry answers with a service description indicating where to locate the service and how to invoke it. The service provider, SySPE, similarly has to generate service descriptions for those services, making its services known by publishing the corresponding service descriptions in a service registry.

The service registry uses all the services descriptions to create a service collection and make it available when services clients request arrive. SySPE validates the service request and sends structured data in an XML file, using the SOAP protocol. The XML file could be validated by the service client using an XSD file.

Zend Framework (ZF) was used to create the web service. ZF is a PHP framework and was used to develop a SySPE application and provide components to work with web services. The components provided by the framework, for web service creation, are mature and well-designed, they offer good integration with the rest of the framework and are comprehensively documented; the entire code is unit-tested and peer-reviewed and there is no licensing fuss around Zend Framework [Ma11]. ZF includes a number of components that enable to work with existing web services as well as create your own. Zend_Service provides a straightforward interface to a number of popular web services like: Amazon, Twitter, Yahoo and Google´s services [Ly09].

A practical example was defined the web service called GET_ INDICATORS, to orient the example in one of the future direction identified by [So13] (import of sustainability indicators) as base of business sustainability performance to upgrade module three. GET_INDICATORS web service allows obtaining all the indicators with their respective

fields stored in SySPE. For it was created a PHP class named Services with one function GET_ INDICATORS. This function establishes the connection with the database, executes the query and return the result in JSON (JavaScript Object Notation) format. JSON is a format for transferring data from one program or system to another and allows making the file with the class names more readable for humans.

ZF allows automatically generate a WSDL XML, based an existing code. To test the WSDL file a PHP class called client was created. The class defines a function WSDL_call and uses a Zend_Soap_Client by pointing a Zend Soap Client instance at the URL returning the WSDL in an XML format to test the service.

The web service is published at Green Service Mall. This component is responsible for the registration of external and internal services offered by IT-for-Green solution to be discovered by the consumers.

7 Web service validation

After the web service design a review and validation of the WSDL is necessary to evaluate the web service. To achieve this, the web services was invoked from "Web Services Validation Tool for WSDL and SOAP 2.1[3]. The Web Services Validation Tool for WSDL and SOAP analyzes and reports problems in SOAP messages and WSDL schemas before implement web services applications. The validation test consisted in visualize the web service input and output to check the desire outcome, the JSON file with all the indicators stored in SySPE application with their respective fields.

The next step was the validation through the SOAP messages. A SOAP message was introduced in the XML file to be validated. The web services GET_INDICATORS was invoke using a "parameter" called "EXAMPLE". The SOAP request is transmitted to the server and the successful SOAP response from the Web Services Validation Tool was received.

This web service allows the communication, interoperability and integration of those platforms through a first example GET_INDICATORS. The web services are a powerful tool to achieve the communication and integration between those systems.

8 Conclusion and Outlook

Nowadays the different business need support of IT resources to monitoring, controlling and supporting decisions making processes as a real imperative in organizations. IT-for-Green and SySPE are two solutions emerged to address these needs. The first covers the complete product life cycle and the and the second allows tracking business

[3] This software was developed and supported by IBM Corporation.

sustainability indicators to help managers to discover which is the overall compliant of sustainability business goals by using CISP, which includes the perception of the different stakeholders.

The communication and integration between those solutions is possible through the web services implementation. To achieve this goal, a web services architecture was defined among services clients: IT-for-Green and SySPE. A web service GET_INDICATORS was defined using a bottom-up model, which make it possible to implement classes first, and then, use a WSDL generating tool to expose methods from these classes as a web service. ZF was used as WSDL automatically generating tool based on existing code.

The future direction of this research is oriented to implement the key functionalities of SySPE, the calculus of Corporate Index of Sustainability Performance and graphic generation as a web service to be used by consumers of IT-for-Green in order to analyze the indicators associated with production process, products or services. Due to practical point of view, this initial integration is a first step to accomplish the goal of integration with SySPE.

9 Acknowledgment

This work is funded by the Eureka SD project (agreement number 2013-2591), that is supported by the Erasmus Mundus program of the European Union.

10 References

[Al04] Alonso, G., et al.: Web Services: Concepts, architectures and applications. Springer Verlag, Berlin Heidelberg, 2004.

[Br13] Bracher, S.: OEPI Platform. In: Dada, A., Stanoevska, K., Gómez, J.M. (eds.) Organizations' Environmental Performance Indicators, 103-116. Springer Berlin Heidelberg, 2013.

[Ca13] Cai, S., et al.: Exploring the role of IT for environmental sustainability in China: An empirical analysis. International Journal of Production Economics 146/2,S. 491-500, 2013.

[Dy02] Dyllick, T., Hockerts, K.: Beyond the business case for corporate sustainability. Business Strategy and the Environment 11/2,S. 130-141, 2002.

[Fi08] Filip, F.G.: Decision support and control for large-scale complex systems. Annual Reviews in Control 32/1,S. 61-70, 2008.

[Gi10] Giesen, N., et al.: ProPlaNET–Web 2.0 based Sustainable Project Planning. EnviroInfo (2010) Integration of Environmental Information in Europe,S. 436-445. Shaker Verlag, 2010.

[HB14] Hack, S., Berg, C.: The Potential of IT for Corporate Sustainability. Sustainability 6/7,S. 4163-4180, 2014.

[He07] Herzig, C., et al.: Nachhaltigkeitsmanagement in Unternehmen: von der Idee zur Praxis: Managementansätze zur Umsetzung von Corporate Social Responsibility und Corporate Sustainability. CSM, Berlin, 2007.

[Hi14] Hilpert, H., et al.: An information system design theory for green information systems for sustainability reporting-integrating theory with evidence from multiple case studies. 22. European Conference on Information Systems (ECIS), Dan Panorama Hotel, Tel Aviv, Israel, 2014.

[It13] IT-for-Green, http://it-for-green.eu/, 12-13-2013

[Ja12] Jamous, N., et al.: Sustainability data integration; A prototypical implementation of a light-weight information system component. In: 26th International Conference on Informatics for Environmental Protection. Shaker Verlag Aachen, 2012.

[LG15] Lippert, S.K., Govindarajulu, C.: Technological, organizational, and environmental antecedents to web services adoption. Communications of the IIMA 6/1,S. 14, 2015.

[Ly09] Lyman, F.: Pro Zend Framework Techniques: Build a CMS Project. Springer Verlag, United States of America, 2009.

[Ma11] Mariën, J.: Zend Framework Web Services A php | architect Guide. Blue Parabola, LLC, Tor ont o, Canada, 2011.

[MA12] Molla, A., Abareshi, A.: Organizational green motivations for information technology: empirical study. Journal of Computer Information Systems 52/3,S. 92-102, 2012.

[Ma13] Mahmoud, T., et al.: Web Service Integration within Next Generation CEMIS. Procedia Technology 9/0,S. 282-290, 2013.

[Me10] Melville, N.P.: Information systems innovation for environmental sustainability. MIS Quarterly 34/1,S. 1-21, 2010.

[Me13] Medel-González, F., et al.: Measuring and Evaluating Business Sustainability: Development and Application of Corporate Index of Sustainability Performance. In: Erechtchoukova, M.G., Khaiter, P.A., Golinska, P. (eds.) Sustainability Appraisal: Quantitative Methods and Mathematical Techniques for Environmental Performance Evaluation, 33-61. Springer Berlin Heidelberg, 2013.

[MMW13] Malhotra, A., Melville, N.P., Watson, R.T.: Spurring Impactful Research on Information Systems for Environmental Sustainability. MIS Quarterly 37/4,S. 1265-1274, 2013.

[Pa09] Padilla, A.: Beginning Zend Framework. In: Anglin, S., Moodie, M. (eds.),S. 387. Springer-Verlag, United States of America, 2009.

[PP09] Petrini, M., Pozzebon, M.: Managing Sustainability with the Support of Business Intelligence Methods and Tools. In: Prasad, S., Routray, S., Khurana, R., Sahni, S. (eds.) Information Systems, Technology and Management, vol. 31, 88-99. Springer Berlin Heidelberg, 2009.

[Ra11] Rapp, B., et al.: IT-for-Green: Next Generation CEMIS for Environmental, Energy and

Resource Management. EnviroInfo 2011: Innovations in Sharing Environmental Observations and Information. Shaker Verlag Aachen, 2011.

[RB13] Rapp, B., Bremer, J.: IT Solutions for EPI Management. In: Dada, A., Stanoevska, K., Marx Gómez, J. (eds.) Organizations' Environmental Performance Indicators, 19-31. Springer, 2013.

[Sc11] Schaltegger, S.: Sustainability Management Control. In: Burritt, R., Schaltegger, S., Bennett, M., Pohjola, T., Csutora, M. (eds.) Environmental Management Accounting and Supply Chain Management, vol. 27, 337-352. Springer Netherlands, 2011.

[So13] Solsbach, A., et al.: Environmental Footprinting in the IT-for-Green Project − A CEMIS Use Case. In: Page, B., Fleischer, A.G., Göbel, J., Wohlgemuth, V. (eds.) 27th International Conference on Informatics for Environmental Protection (EnviroInfo 2013),S. 318-326. Shaker Verlag, Hamburg, 2013.

[St05] Steurer, R., et al.: Corporations, stakeholders and sustainable development I: a theoretical exploration of business–society relations. Journal of Business Ethics 61/3,S. 263-281, 2005.

[TP13] Tachizawa, T., Pozo, H.: Monitoramento do passivo socioambiental com o suporte da tecnologia de informação. JISTEM J.Inf.Syst. Technol. Manag. (Online) 7/1,S. 95-120 2010.

[Wo08] Wohlgemuth, V.: Berichte aus der Umweltinformatik. Konzepte, Anwendungen, Realisierungen und Entwicklungstendenzen betrieblicher Umweltinformationssysteme (BUIS). Workshop 1. Berliner BUIS-Tage, Aachen: Shaker, 2008.

[Xi05] Xiaoying, B., et al.: WSDL-based automatic test case generation for Web services testing. In: Service-Oriented System Engineering, 2005. SOSE 2005. IEEE International Workshop, S. 207-212. 2005.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Big Data Inspired Water Management Platform for Sensor Data

Viktor Dmitriyev[1], Jorge Marx Gómez[2] and Manuel Osmers[3]

Abstract: The usage of the data generated by the various types of sensors is addressed to be one of the promising source of improvements and one of the core stones of the Industry 4.0 initiative of the German government. The domain of sensor data management continuously demonstrates it's ability to improve currently existing processes in various directions. Despite theoretically applicability of the same approach of the sensor data management in the domains other that water, current work focuses on demonstrating preliminary results of the research done with focus on water management. The preliminary results demonstrated in the form of principal platform architecture. The main idea behind the paper is to show the principal architecture in order to open a future opportunity for building a prototypical implementation of the proposed architecture later on. The principal architecture has an big data phenomena principals like volume, velocity and variety by-design.

Keywords: sensors, smart data, big data, water management, internet of things

1 Introduction

Water resource is a starting point of live for all living species on our planet. In case of the space missions in searching extraterrestrial life, the "target number one" is searching for availability of water. Most life on earth need freshwater, but only 0.3 % of the water is on the planet calculated as drinkable [Gl93]. So it is important to use our water sustainable and sensible. But despite there are a lot of discussions about the water all around the globe, the main question that still needs an answer is "How we can support water sustainable usage with information and communications technologies (ICT)?" Despite the topic of resources sustainability is not a new one, the water as most important resource for human, nature and industry needs more attention from different viewpoints, an example of such is ICT. Water is no longer an endless resource; rather we have to think about dependable methods and look for approached to enhance sustainability aspects of water utilization.

It is a common and widespread idea among people living in Germany that they have no problems with the water supply, but within the next few years people with high probability may face some of the problems such as other countries are facing.

[1] University of Oldenburg, Business Information Systems / VLBA, Ammerländer Heerstr. 114-118, 26129 Oldenburg, viktor.dmitriyev@uni-oldenburg.de
[2] University of Oldenburg, Business Information Systems / VLBA, Ammerländer Heerstr. 114-118, 26129 Oldenburg, jorge.marx.gomez@uni-oldenburg.de
[3] University of Oldenburg, Business Information Systems / VLBA, Ammerländer Heerstr. 114-118, 26129 Oldenburg, manuel.osmers@uni-oldenburg.de

Especially, the nitrate concentration in the water of the North West region in Germany is a promising problem, next to salinization and population growing, which will defiantly attracting far more attention in the future. Some institutions are working on the topic of prevention such problems. They started to use information and communications technologies to visualize and compute the scenarios, in which they simulate the problems that are about to appear. By conducting different simulations and visualizations they are trying searching for applicable solutions.

The other side of the promising ICT utilization within the context resource management, and water management in particular, is the usage of other data sources like Open Data, as well as services and interfaces developed on top of the Open Data [Da10, MRS13, Gu11]. The Open Data is emerging as one of the most promising value adding data-centric approaches. Open Data is attracting more and more attention from the business and government sides. In 2011 McKinsey predicted that big data market for the European Sector by 2020 is going to be around €250 billion per year [Ma11], where the Open Data is playing a really great role. The further steps for the work are to combine Open Data with proprietary data from water suppliers and state authorities in order to create more sustainable and value adding processes heavily supported by decisions made upon data.

Especially, the concepts of openness within the Open Data approach is promising to make cities ”smarter”. An idea of making cities smarter attracts more attention to the ”Smart Cities”[To13] phenomena. Following ideas and concepts behind Smart Cities allows resource such as urban city services, their infrastructure utilization to being more sustainable, which is very important, even crucial, with the continuously growing urbanization [Ki14, Ha12]. Despite usage of the water is very common in almost any locations, the possibility of proper and most optimal water resource utilization without much investments (with just ICT driver innovations) is very promising in highly populated and well organized locations such as cities.

2 Main Idea

The usage of the data generated by the various types of sensors, which can be a part of the day-to-day used smartphone device, or some other dedicated devices that is able to measure temperature, or gas pressure within the car engine, are agreed to be one of the promising source of improvements and one of the core stones in the Internet of Things (IoT) [AIM10, Gu13] and Industry 4.0 initiative of the German Government [Ka13]. The domain of sensor data management continuously demonstrates it’s ability to improve the current existing processes in various directions.The current work is focused on demonstrating preliminary results of the research done with focus on water management. The idea of the work is first of all to show the principal architecture and than build a prototypical implementation of the proposed architecture later on. The principal architecture has an big data principals like volume, velocity and variety. Current work is going to demonstrate the principal architecture that will enlarge the potential of water management, in particular, the architecture is addressing following requirements (a) consolidate sensor and event

data, (b) store data, (c) offer interactive analysis and standard reporting, (d) offer pluggable components approach.

In the water management there are number of different measured rates that can be traced by particular sensor (e.g. pressure, fluid-flow, level of water pollution, salt water intrusion, etc.). By adding on top of the sensor data, a business data from the companies and the data from customers a very interesting scenarios for the water management can be create by the experts from various domains (e.g. sustainability). Such huge amounts on data from many different sources creates a 3 mostly common dimension of the big data: Volume, Velocity and Variety [La01]. On the other hand, the approach is designed for usage of business and customer data for economic and sustainable goals as well, for example by incorporating them together into prediction models or sustainable reporting. Due to the above mentioned the work has a relevance to the Environmental Management Information System (EMIS). An implementation of the use cases with EMIS inside proposed architecture should demonstrate the applicability of ideas described by the current work on more granular level.

The idea is to demonstrate abilities of ICT to help in solving many types of problems, which water management is currently facing and going to face in the future. Taking into consideration addressed architecture will help the organization to be more flexible in terms of data storage and processing, as well as prepare the infrastructure, which is mainly focused on the data analysis part, for future changes and generation of enormous amounts of sensors data. The general approach proposed by the work to meet addressed requirements can be divided into following steps: (a) consolidate historical water related data; (b) consolidate real time water data from sensors; (c) visualize water reservoirs with respect to their parameters volume, geo-location, water consumption and water recovery potential; (d) apply statistical methods (e.g., data mining, machine learning, etc.).

3 Architecture Overview

General overview of the architecture of the Water Management Platform for Sensor Data is shown on the Fig.1. The core component of the overall architecture is an In-Memory Database, which can be defined as an database management system (DBMS) that utilizes capacity of main memory and uses disks as the secondary storage [PZ12]. Usage of main memory as the primary data storage allows to overcome physical obstacle of data storages based on the magnetic rotating disks. The usage of In-Memory Database will significantly simplifications data model (because most of the works around such as pre-aggregations within the analytical cubes are not more required), it also should be bring a significant performance boots to the all types of queries (reporting, add-hoc, OLAP, etc.).

The primary roles of the Water Management Platform for Sensor Data is to offer centralized data storage, integration and analysis facilities for each particular water related scenario. Part of the platform, that is responsible for the data storage consists out of 5 main components and additional ability of plugging custom component: (1) Event Processing Engine; (2) Long-Term Storage; (3) In-Memory Database; (4) Reporting; (5) Interactive

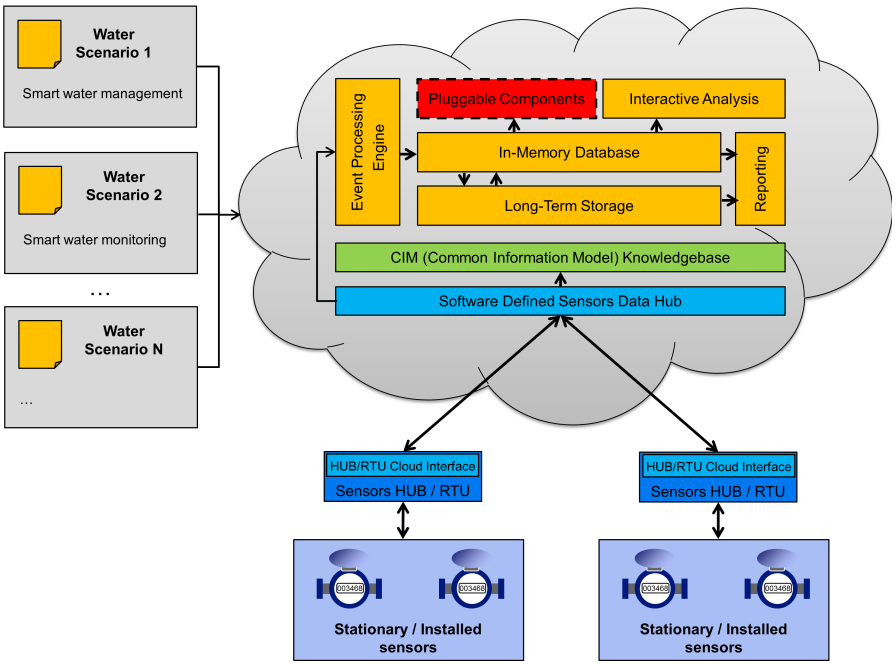


Fig. 1: Architecture Of Water Management Platform for Sensor Data

Analysis. The platform gets sensor data through the Software Driven Sensor Data Hub. The internal communications between components within the platform are facilitated by commonly acceptable standards and protocols (OData, ODBC, HTTP/S, etc.). In the particular case, communications between Software Driven Sensor Data Hub and Sensors/RTU-s will be based on CIM (Common Information Model) that brings interoperability to the target architecture and also support computability with a wider range of devices.

As it's shown on the Fig.1 software components are run inside the private cloud, such approach allows to scale the overall architecture with accordance of demand. Using enhanced cloud solution stack will also prepare a future landscape for the deployment of the targeted solution. Integration of advanced sensors, such as RTUs (Remote Terminal Unit), in the architecture through the Sensor Data Hub, as it's shown in Fig. 1, will bring flexibility to the architecture in the sense of integrating new types of RTUs (existing on the marked and future ones).

As it was mentioned above, every component of the platform should run in the cloud stack. The Event Processing Engine can be implemented with the help of framework that allows processing of continuous data, for example Odysseus Framework [Ap12]. The Long-Term Storage cane be implemented as a distributed database or as a distributed file system with processing engine on top. In particular, the role of the Long-Term Storage can be overtaken by e.g. Apache Hadoop and other components from it's Ecosystem [Va13]. An In-Memory Database will be used for the use case, in which requirements are demanding data process-

ing heavy operations such as ad-hoc queries, KPI visualization or even data mining and machine learning. Interactive Analysis of data will integrate differed analytical capabilities such as highly customizable analytical views. The predictive analytics can be as well integrated into the platform through Plugable Component. Reporting can be done by means of any openly available tool such, for example Pentaho BI [LBF14] or even some other in-house developed software solution.

4 Related Works

There are number of projects contributing to the sustainability of the particular city and making a city "a smart" one by integrating various of data streams within a single system and processing integrated data in order to achieve better level of understating of ongoing events and processes within the particular environment and context. One of the such projects is the CityPulse [Ko14, Ga15], the project that was funded by the EU and was conducted by the multinational consortium. The major goal of the CityPulse project is to benefit from interconnected ecosystem of the city and improve sustainability by more efficient (optimal) operations performed with city services and infrastructure. CityPulse tried [Ko14, Ga15] to integrate various source of the streaming data generated by the devices together with social data streams for in a about city into an single large scale real-time system.

The other example of the related works is the project called CyberWater [CMI13]. The main goal of the project is to conduct a system, which is able provide a decision support in risk related environments focusing on an integrated management of watersheds and pollution caused by some accidents. Due to the overwhelming amounts of data and geographical heterogeneity of the data sources (sensors), the core system is seen to be a distributed and based on the heavily based on the Service-Oriented Architecture (SOA) principles for interoperability.

Despite there are a list of similarities between the proposed Water Management Platform for Sensor Data and CyberWater in main goals, there are a list of principal differences between them on architecture level. The main goal of the CyberWater is to help in decisions, whether the main goal of the proposed approach is to simplify working with data on all levels and give a strong and robust tool to the domain experts to work with data without investing much time into establishing a proper data engineering pipeline. Another difference to be mentioned is an interaction between components in proposed approach. A particular interaction is done through previously defined protocols, which works with significant efficiency in comparison with Service-Oriented Architecture (SOA), because the decision whether to use or not to use particular protocol for connections, as well as depth of integration into protocol of an component, can be done based on various requirements (performance, fault-tolerance. etc.). Such elasticity in protocols integration make proposed architecture flexible. Other significant difference is that the Water Management Platform for Sensor Data is user case independent. One of the main focus of the proposed approach is to make a transfer process of every single data point, which is generate by sensors, transparent for the end users.

5 Conclusions and Future Works

Current work demonstrates demand for the platform, which is not only able to manage water related data, but is also able to face upcoming challenges addressed by fast evolving initiatives such as Industry 4.0 and Internet of Things. The general overview of the principal architecture for the Water Management Platform for Sensor Data was shown and described. Future work will be the detailed description of the water use cases and the prototypical implementation of the proposed platform together with an industry partner, which are showing a great interest in conducting a prototype based on the proposed architecture and foreseeing a huge benefit for themselves and their customers.

References

[AIM10] Atzori, Luigi; Iera, Antonio; Morabito, Giacomo: The internet of things: A survey. Computer networks, 54(15):2787–2805, 2010.

[Ap12] Appelrath, H.-Jürgen; Geesen, Dennis; Grawunder, Marco; Michelsen, Timo; Nicklas, Daniela: Odysseus: A Highly Customizable Framework for Creating Efficient Event Stream Management Systems. In: Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems. DEBS '12, ACM, New York, NY, USA, pp. 367–368, 2012.

[CMI13] Ciolofan, Sorin N; Mocanu, Mariana; Ionita, Anca: Distributed Cyberinfrastructure for Decision Support in Risk Related Environments. In: Parallel and Distributed Computing (ISPDC), 2013 IEEE 12th International Symposium on. IEEE, pp. 109–115, 2013.

[Da10] Davies, Tim: Open data, democracy and public sector reform. A look at open government data use from data. gov. uk, 2010.

[Ga15] Ganz, Frieder; Puschmann, Daniel; Barnaghi, Payam; Carrez, Francois: A Practical Evaluation of Information Processing and Abstraction Techniques for the Internet of Things. 2015.

[Gl93] Gleick, Peter H et al.: Water in crisis: a guide to the world's fresh water resources. Oxford University Press, Inc., 1993.

[Gu11] Gurstein, Michael B: Open data: Empowering the empowered or effective data use for everyone? First Monday, 16(2), 2011.

[Gu13] Gubbi, Jayavardhana; Buyya, Rajkumar; Marusic, Slaven; Palaniswami, Marimuthu: Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems, 29(7):1645–1660, 2013.

[Ha12] Hancke, Gerhard P; Hancke Jr, Gerhard P et al.: The role of advanced sensing in smart cities. Sensors, 13(1):393–425, 2012.

[Ka13] Kagermann, Henning; Helbig, Johannes; Hellinger, Ariane; Wahlster, Wolfgang: Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry; Final Report of the Industrie 4.0 Working Group. Forschungsunion, 2013.

[Ki14] Kitchin, Rob: The real-time city? Big data and smart urbanism. GeoJournal, 79(1):1–14, 2014.

[Ko14] Kolozali, Sefki; Bermudez-Edo, Maria; Puschmann, Daniel; Ganz, Frieder; Barnaghi, Payam: A Knowledge-based Approach for Real-Time IoT Data Stream Annotation and Processing. In: Internet of Things (iThings), 2014 IEEE International Conference on, and Green Computing and Communications (GreenCom), IEEE and Cyber, Physical and Social Computing (CPSCom), IEEE. IEEE, pp. 215–222, 2014.

[La01] Laney, Douglas: 3-D Data Management: Controlling Data Volume, Velocity and Variety. META Group Research Note, February, 6, 2001.

[LBF14] Lapa, Joaquim; Bernardino, Jorge; Figueiredo, Ana: A Comparative Analysis of Open Source Business Intelligence Platforms. In: Proceedings of the International Conference on Information Systems and Design of Communication. ISDOC '14, ACM, New York, NY, USA, pp. 86–92, 2014.

[Ma11] Manyika, James; Chui, Michael; Brown, Brad; Bughin, Jacques; Dobbs, Richard; Roxburgh, Charles; Byers, Angela Hung; Institute, McKinsey Global: Big data: The next frontier for innovation, competition, and productivity. 2011.

[MRS13] Magalhaes, Gustavo; Roseira, Catarina; Strover, Sharon: Open Government Data Intermediaries: A Terminology Framework. In: Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance. ICEGOV '13, ACM, New York, NY, USA, pp. 330–333, 2013.

[PZ12] Plattner, Hasso; Zeier, Alexander: In-Memory data management: Technology and applications. Springer Science & Business Media, 2012.

[To13] Townsend, Anthony M: Smart cities: big data, civic hackers, and the quest for a new utopia. WW Norton & Company, 2013.

[Va13] Vavilapalli, Vinod Kumar; Murthy, Arun C; Douglas, Chris; Agarwal, Sharad; Konar, Mahadev; Evans, Robert; Graves, Thomas; Lowe, Jason; Shah, Hitesh; Seth, Siddharth et al.: Apache hadoop yarn: Yet another resource negotiator. In: Proceedings of the 4th annual Symposium on Cloud Computing. ACM, p. 5, 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIC 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Modeling for sustainable product development strategies with general morphological analysis

Zhongkai Li[1], Jorge Marx Gómez[2]

Abstract: Sustainable product development is currently of concern by scientists within both academia and engineering. Despite existing literature proposed related to detailed aspects for design for sustainability (D4S), the modeling of life cycle driving forces for D4S in different levels remains insufficient. A morphological analysis (MA) approach was applied to explore the strategies in sustainable product development in this work. With the experts' brainstorming and MA framework, the results for solutions to D4S were obtained under the support program from the Swedish Morphological Society. The specific factors and states for six sustainable development scenarios were acquired and compared. Platform-based innovation, flexible manufacturing system, multi-material lightweight design, energy efficiency priority, disassemble to reuse, remanufacturing, recycling (3R), LCA and CAX software are identified as key strategies for D4S. The sustainability-improving strategies for companies with an eco-design basis, small to medium enterprises (SME), and governments are also briefly discussed respectively.

Keywords: Design for sustainability, D4S, Eco-friendly product, product development strategy, morphological analysis

1 Introduction

Product design for sustainability (D4S) [CD09] is currently highlighted by scientists and industrial engineers to obtain ecological (eco-) friendly products. The sustainable products can be defined as products that provide environmental, social, and economic benefits, while protecting health and welfare as well as maintaining the environment over their full life cycle from raw materials extraction and use, to eventual disposal and reuse [GW15]. D4S is a general destination for electromechanical product design. To achieving this, several detailed aspects including design for environment (DFE) [IG10], design for life cycle (DFLC) [RR10], design for disassembly (DFD) [BO07], design for recycling (DFR) [UF13], and material selection (MS) in eco-design [ST13], have been researched in academia. In this study, focus is placed on physical products, such as automobiles, engineering machinery, and household electronics. The goal of this focus is to explore the main factors and important states related to sustainable product development. The findings will help to improve government policy and enterprise practice to promote the research and development of eco-friendly products.

[1] School of Mechatronics Engineering, China University of Mining & Technology, 221116 Xuzhou, China, zhongkai.li@uni-oldenburg.de
[2] Department of Informatics, Very Large Business Applications(VLBA), University of Oldenburg, 26129 Oldenburg, Germany, Jorge.Marx.Gomez@uni-oldenburg.de

Limited studies have been published about the criteria, policy, or factors for sustainable product design. For instance, Maxwell *et al.* [MV03] obtained the criteria for optimising sustainability through process research of sustainable products and services development (SPSD). The criteria included functionality, environmental impacts, social impacts, economic impacts, quality, market demand, customer requirements, technical feasibility, and compliance with legislation and industry/technical specifications. However, the volume of criteria poses a challenge and a trade-off might be needed in specific SPSD environments. Zhang *et al.* [ZX12] presented a theoretical model for new product development by investigating the interactions among customers' preference, firms' product strategies, and government subsidy policies. They found to motivate firms to choose eco-friendly product design strategies, government should put forward effective subsidy policies. Tu *et al.* [TC13] used the survey and data analysis method to analyze the impact factors and strategy of sustainable product development under Corporate Social Responsibility (CSR) in Taiwan. Results from questionnaire interviews were analyzed by using frequency distribution, factor analysis and content analysis approaches. The internal and external major factors were then obtained to guide further sustainable product development. One common feature of the above three studies is that they identified the parameters to evaluate or assist product sustainability. A shortcoming of these pioneering studies lies in the fact that the product lifecycle phases are overlooked.

Morphological analysis (MA) and box method is applied in this work to model the reasons and results for product D4S. General morphological analysis (GMA, briefly named as MA) [RS02] is a method for structuring and investigating the total set of relationships contained in multi-dimensional, non-quantifiable, problem spaces. MA was initially proposed by Fritz Zwicky – the Swiss astro-physicist and aerospace scientist based at the California Institute of Technology [ZW69]. MA was later further developed by Tom Ritchey in the Swedish Morphological Society. MA is suitable for mapping the relationships within multiple factors and related distinct states according to expert experience. An alternate definition suggests that MA is a computational framework to transform the consistent data within elements given by experts into a visualized model showing reasons and results. The MA acquired model serves as an experimental platform to analyze and compare different operational strategies. A computer-aided MA system was developed by the Swedish MA society and over one hundred projects have been implemented under its instructions. Examples that apply this technique using this software include modeling complex socio-technical systems [RI02], modeling multi-hazard disaster reduction strategies [RI06], and threat analysis for the transport of radioactive material [RI07].

Considering the product conception, raw material, production process, distribution, consumption, and end-of-life phases within a product life cycle, eight factors are selected and multiple distinct states for each factor are defined. The factors cover the design, manufacturing, material selection, energy efficiency, EOL recycling, eco-impact analysis, software tools, and design scenarios with their different states as much as possible. Through the assessment data from experts in different disciplines and computer-aided

MA, drivers for different levels of sustainable product design were discovered. Comparisons for multiple scenarios and analysis for promoting sustainable design within specific scenarios are also discussed.

Section 2 of this paper describes the MA approach and definition data by experts brainstorming. Section 3 outlines the findings from the model via MA. Section 4 discusses the results and compares several sustainability improving strategies. Section 5 provides conclusions of the research.

2 Methodology

The computer-aided MA approach to analyze the reasons for different sustainable development scenarios has been used. The framework of MA and definition data for sustainable product development strategies are described as follows.

2.1 Framework for MA

MA is an easy-to-use computational framework to find many compatible results within multiple reasons. The function of MA is to explore solutions for qualitative, uncertain, and complex problems without mathematical models. Thus, MA is a typical human-computer interaction process with several iterations. As showed in Fig. 1, the internal MA process and the external support from human experts and software tools are illustrated. MA contains three main steps [RS02] as the bold numbers represented in this figure:

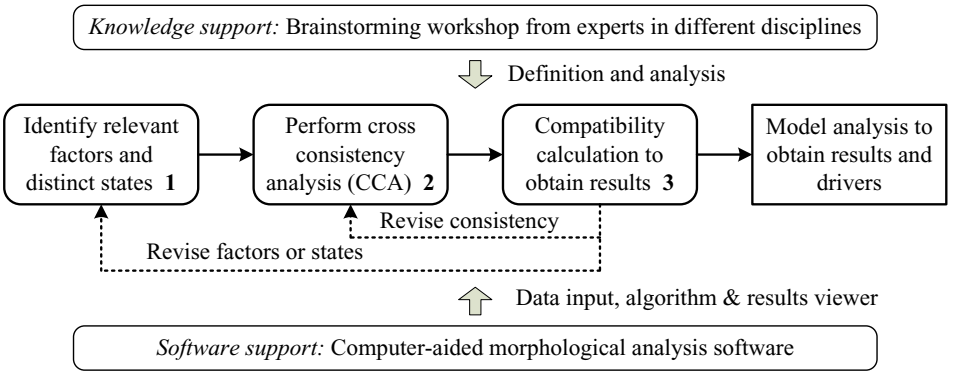


Fig. 1: Framework and external support for MA

Step 1: Setting the factors and distinct states for the problem. In another word, setting the parameters and different values for the strategic analysis problem. The factors are desired to cover different aspects of the problem as completely as possible with a maximum of eight. The states within one factor are better to be mutually independent,

namely, only one may be selected. An average of five to six states per factor were defined.

Step 2: Input the cross-consistency analysis (CCA) matrix to reduce the feasible solution space. This step is to represent the inherent relations between pairs of states within different factors. Three values such as total consistency, paritcal consistency, and no consistency can be represented with " -", " K", and " X", respectively.

Step 3: Analyzing the strategies with the obtained model under the environment of MA software. The concerned results and reasons can be acquired and compared to find interesting points.

Note that the experts are better to be in different disciplines to involve multiple knowledge backgrounds. The second step is most time-consuming because of the large amount of assessment data via the brain-storming workshop of experts. Note that the interation process within the three steps is completely normal, and designers are learning more about the nature of their problem space. This article doesn't discuss the detailed MA algorithm, and the focus lies in findings on product ecological development strategies. Under the research support program from the Swedish Morphological Society [SW15], the CCA matrix was supplied in order to obtain the visualized results by MA software also with several iterations.

2.2 Problem definitions

Since the aim of the research is to explore the drivers for different levels of product development scenarios, eight factors are selected as showed in Table 1. The table contains seven kinds of drivers and one kind of result. Considering design for life cycle principles, the *design innovation types*, *manufacturing modes in enterprise*, *material selection approaches*, *energy efficiency levels in usage*, *recycling levels in end-of-life (EOL)*, *eco-impact analysis modes*, and *computer-aided software tools* are selected as the main factors for the scenarios in D4S. Furthermore, five to six distinct states are provided under each factor according to the state-of-the-art techniques in enterprises with different levels. The detailed states in each factor are also listed in Table 1. To ease the presentation in context, each factor and states are named with letters and numbers.

Table 1: Factors and related states for product development scenarios

Factors	States
A. Design innovation type	A1. Totally innovative design
	A2. Platform-based innovation
	A3. Appearance innovation
	A4. Design for individualized order
	A5. No innovation
B. Manufacturing mode	B1. Flexible manufacturing system
	B2. Multiple manufacturing cell
	B3. Mass production line

	B4. Distributed machine tools and low level assembly
	B5. Small amount customization
C. Material selection	C1. Multiple material lightweight design
	C2. Only use primary raw materials
	C3. Partially use recycled raw materials
	C4. Cost and safety priority based material composition
	C5. Material selection without toxicity
	C6. Case-based material selection without innovation
D. Energy efficiency in service	D1. Energy efficiency priority
	D2. Cost priority, & energy followed
	D3. Without considering energy efficiency, only cost priority
	D4. Acquiring energy efficiency by calculation
	D5. Acquiring energy efficiency by physical experiment
E. Disassembly & recycling for EOL products	E1. Disassemble to reuse, remanufacturing, and recycle
	E2. Cascade use with DFD/DFR
	E3. Partial recycle, & partial landfill
	E4. Only recycle as raw materials
	E5. Only energy based recycle
	E6. All disposal in landfill
F. Eco-impact analysis	F1. LCA
	F2. Only material eco-impact analysis
	F3. Only manufacturing resource efficiency analysis
	F4. Qualitative comparison for different products
	F5. Multi-attribute decision making by experts
	F6. No eco-impact analysis
G. Software tools	G1. 2D CAD drawing
	G2. 3D CAD modeling and drawing
	G3. CAD/CAM/PLM
	G4. CAD, CAE and analysis tools without LCA
	G5. CAX/LCA
S. Scenarios in design for sustainability	S1. Eco-products with good sustainability
	S2. Products with extrem complexity and high reliability
	S3. Products with cost priority and satisfying eco-regulations
	S4. Products with cost priority and functional satisfaction with little eco-involved
	S5. Low cost and low quality products without eco-involved
	S6. Only functional prototypes

Six scenarios in D4S were defined with the highest one as eco-products with good sustainability to the lowest one as only functional prototypes. The sustainability level from S1 to S6 is decreases in priority. Note that the A1 - totally innovative design means there has principle innovation and little reference case. E2 – cascade use with DFD/DFR represents the EOL products are disassembled and recycled as other kinds of secondary products, with little reuse or remanufacture to original ones. F3 – only manufacturing resource efficiency analysis is to calculate the ratio of material output and input during processes, which likes the financial cost analysis under Enterprise Resource Planning (ERP) system. Most of the states within factors are independent. Although we tried to keep the states within one factor to be independent as much as possible, it cannot be

guaranteed that all states are distinct because of the coupling in design process. For instance, the D1 and D4 can be coexisted because with plenty of design data and model, the energy efficiency can be calculated (D4) within an energy efficiency priority design scenario (D1).

The cross-consistency of the elements are defined by the experts brainstorming workshop. Because of the limit of the context, a partial CCA matrix is shown in Fig. 2. The following key to annotate the assessments is used:

- "-" (hyphen) = Good fit, or best fit, or optimal pair.
- "K" = Possible, could work, but not optimal.
- "X" = Impossible or unrecommended idea.

Because the CCA is a two-dimensional and undirectional matrix, it is defined that the columns of the CCA do not contain the last factor, and the first factor is also deleted from the rows in CCA. Note that one must not have a full column or a full row of only "X"s in a parameter-block. There should be at least one "- " in each column and row for each parameter-block. An assessment for any of the cells in the CCA matrix is required.

design-step2-R3.scn

		Design innovatio					Manufacturing m					Material selection						Energy efficiency				Disass	
		Totally innovative design	Platform-based innovation	Appearance innovation	sign for individualized order	No innovation	exible manufacturing system	Multiple manufacturing cell	Mass production line	ools and low level assembly	Small amount customization	a material lightweight design	ly use primary raw materials	y use recycled raw materials	based material composition	rial selection without toxicity	selection without innovation	Energy efficiency priority	st priority, & energy followed	efficiency, only cost priority	ergy efficiency by calculation	ency by physical experiment	emanufacturing, and recycle	Cascade use with DFD/DFR
Disassembly & recycling for EOL products	Disassemble to reuse, remanufacturing, and recycle	-	-	X	X	X	-	X	X	X	X	-	X	-	X	-	X	-	K	X	-	X	
	Cascade use with DFD/DFR	-	-	K	-	X	-	-	X	X	X	-	K	-	-	-	X	-	-	K	-	X	
	Partial recycle, & partial landfil	X	-	-	-	X	K	-	-	-	K	K	-	K	-	-	K	X	K	-	X	-	
	Only recyle as raw materials	X	-	-	-	X	K	-	-	-	X	K	-	-	-	K	X	K	-	-	K	K	
	Only energy-based recycle	X	X	-	X	-	X	X	X	-	K	X	K	-	K	X	-	X	X	-	X	-	
	All disposal in landfill	X	X	X	K	-	X	X	X	K	-	X	X	X	X	-	-	X	X	-	K	K	
Eco-impact analysis	LCA	-	-	X	X	X	-	-	X	X	X	-	K	-	X	-	X	-	X	X	-	X	-	-
	Only material eco-impact analysis	X	-	-	-	X	X	-	-	X	X	-	-	-	-	-	X	-	-	-	-	K	K	-
	Only manufacturing resource efficiency analysis	X	K	-	K	X	-	K	K	X	X	X	X	X	-	X	X	-	K	X	X	-	-	-
	Qualitative comparison for diff	X	X	K	X	-	X	X	K	-	X	X	-	X	-	X	-	X	X	-	K	-	X	X

Fig. 2: Partial cross consistency matrix inputed in morphological analysis software

3 Findings

With the predefined data, the model for sustainable product development scenarios is obtained and visualized by MA software. The model is regarded as an input-output

device. As showed in Fig. 3, if one element in the matrix is clicked to be shown in red color, related reasons for this scenario will be highlighted in blue to form a mapping between drivers and results. If one element can not find any drivers in blue, there are no reasons for this scenario under the current problem definitions.

For achieving eco-products with good sustainability (S1), totally innovative design (A1) or platform-based innovation (A2) can be used. Also, flexible manufacturing system (B1) with multi-function robotics and advanced numerical control machine tools will support the eco-manufacturing process. In material selection, the multiple material lightweight design (C1), paritially use recycled raw materials (C3), and material selection without toxicity (C5) are all preferred. The products are desired to be energy efficiency priority (D1) and acquiring efficiency rate be calculation (D4). Further, the products should be designed with ease of disassemble to reuse, remanufacturing, and recycle (3R), or they can be easily cascade use with DFD/DFR. Moreover, CAX/LCA software tools (G5) should be implemented in the development process to support quantitative life cycle analysis (F1) for effective eco-impact assessment.

Design innovation type	Manufacturing mode	Material selection	Energy efficiency in service	Disassembly & recycling for EOL products	Eco-impact analysis	Software tools	Scenarios in design for sustainability
Totally innovative design A1	Flexible manufacturing system B1	Multiple material lightweight design C1	Energy efficiency priority D1	Disassemble to reuse, remanufacturing and recycle E1	LCA F1	2D CAD drawing	Eco-products with good sustainability S1
Platform-based innovation A2	Multiple manufacturing cell	Only use primary raw materials	Cost priority, & energy followed	Cascade use with DFD/DFR E2	Only material eco-impact analysis	3D CAD modeling and drawing	Products with extrem complexity and high reliability
Appearance innovation	Mass production line	Partially use recycled raw materials C3	Without considering energy efficiency, only cost priority	Partial recycle, & partial landfill	Only manufacturing resource efficiency analysis	CAD/CAM/PLM	Products with cost priority and satisfying eco-regulations
Design for individualized order	Distributed machine tools and low level assembly	Cost and safety priority based material composition	Acquiring energy efficiency by calculation D4	Only recyle as raw materials	Qualitative comparison for different products	CAD, CAE and analysis tools without LCA	Products with cost priority and functional satisfaction with little eco-involved
No innovation	Small amount customization	Material selection without toxicity C5	Acquiring energy efficiency by physical experiment	Only energy-based recycle	Multi-attribute decision making by experts	CAX/LCA G5	Low cost and low quality products without eco-involved
		Case-based material selection without innovation		All disposal in landfill	No eco-impact analysis		Only functional prototypes

Fig. 3: Drivers for eco-products with good sustainability

Table 2: Reasons for different sustainable product development scenarios

Scenarios	A	B	C	D	E	F	G
S1	A1; A2	B1	C1; C3; C5	D1; D4	E1; E2	F1	G5
S2	A2; A4	B1; B2	C1; C3; C5	D1; D4	E2	F2; F5	G3; G4
S3	A2; A3; A4	B2; B3	C2; C3; C5	D2; D5	E3; E4	F2	G2; G3
S4	A3	B3; B4	C2; C4	D3; D5	E4	F4; F5	G2; G4
S5	A3; A5	B4	C4; C6	D3	E5	F4; F6	G1; G2
S6	A4; A5	B4; B5	C6	D3	E6	F6	G1

For ease of showing and comparison, the reasons for different design scenarios are listed in Table 2. The products with extreme complexity and high reliability (S2) are special for complex electromechanical equipment. Because of its high reliability and specilized

requirements, the drivers for it are platform-based innovation (A2) or design for order (A4) with flexible manufacturing system (B1) or multiple manufacturing cells (B2). The states in material selection and energy efficiency are same as S1. Moreover, it desires cascade use with DFD/DFR (E2) with only material eco-impact analysis (F2) or multi-attribute decision making by experts (F5). The suitable softwares are CAD/CAM/PLM (G3) or CAD, CAE and analysis tools without LCA (G4).

For products with cost priority and satisfying eco-regulations (S3), the drivers in design modes are platform-based innovation (A2), appearance innovation (A3), or design for individualized order (A4). The multiple manufacturing cell (B2) or mass production line (B3) are both allowed with only use primary raw materials (C2), partially use recycled raw materials (C3), and to select material without toxicity (C5). Other preferred drivers for this scenario are cost priority & energy followed (D2) or acquiring energy efficiency by physical experiment (D5). EOL strategies are paritial recyle & paritial landfill (E3) or only recycle as raw materials (E4). Only material eco-impact analysis (F2) with 3D CAD modeling (G2) or CAD/CAM/PLM integration (G3) are also driving forces.

The above three scenarios with good product sustainability are our most preferred ones. The following three results from S4 to S6 only exhibit the product development conditions for some lower level enterprises or plants. Products with cost priority and functional satisfaction with little eco-involved (S4) is stilled applied by large number of companies. Their common points are appearance innovation (A3) with mass production line (B3) or distributed machine tools & low level assembly (B4). They perform cost and safety priority based material composition (C4) or only use primary raw materials (C2). They are just considering cost priority (D3) and acquiring energy efficiency by physical experiment (D5). The EOL products can only be recycled as raw materials (E4). Also, lower level eco-impact such as qualitative comparison for different products (F4) or multi-attribute decision making by experts (F5) are still applied. The 3D CAD (G2) or CAD, CAE and analysis tools without LCA (G5) are implemented. The drivers for S5 and S6 can also be found in Table 2 and but not further discussed.

4 Discussion

Several key states in multiple factors are found for sustainable product development. We obtain that the platform-based innovation (A2), flexible manufacturing system (B1), multi-material lightweight design (C1), energy efficiency priority (D1), disassemble to 3R (E1), CAX (G5), and LCA (F1) are all important strategies for achieving eco-friendly products with sustainability. Though totally innovative design (A1) can involve principle innovations in some degree. But it is extremely difficult to develop totally new products "without former cases" successfully. Moreover, design for individualized order (A4) is always done by reconfiguring from an original kernel product. Thus, platform-based innovation (A2) should be carefully focussed by R&D managers to develop eco-friendly and customized products.

Increasing products' sustainability level from the current eco-design level is concerned. Revising the energy efficiency or materials are now widely utilized by companies to reduce the eco-impacts of their outcomes. For example, a household washing machine is redesigned to be better in energy efficiency and use lightweight materials with modularization in structure to ease maintenance and EOL recycling. But the resource efficiency in manufacturing and quantitative LCA are still lacking behind because of their large financial investment. Thus, the LCA and flexible manufacturing system are desired to be concentrated by groups with a degree of basis on eco-design.

There are still suggestions for design scenarios which mainly considering cost priority and functional satisfaction with little or none eco-involved. It is recognized that only providing low level products will be not benefit for the sustainable development of the SMEs. To improve product design in materials, energy efficiency, and to update the manufacturing facilities with flexible multi-functional cells are suitable strategies for SMEs. Further, the market access standard, producer responsibility, and recycling network for EOL products should be more focussed by governments and society.

5 Conclusions

Through the sustainable product development strategy analysis by MA, six key strategies within multiple lifecycle phases are identified for providing eco-friendly products. Specific drivers are found for different sustainable design scenarios. These drivers are regarded as strategies that can be considered by enterprises or governments to promote eco-development, rather than some detailed techniques that can be directly utilized. The platform-based design innovation, flexible manufacturing system, energy efficiency priority, design for recycling, and LCA, CAX software are desired to be focused and integrated. They will be benefit to achieve sustainable product development in higher level. Future work lies in the software architecture research for promoting product D4S.

Acknowledgment

This work is funded by the National Natural Science Foundation of China under grant 51475459, and Fundamental Research Funds for the Central Universities of China under Grant 2012QNB15. We gratefully acknowledge Prof. Tom Ritchey and the support program from Swedish Morphological Society. Sincere appreciation is extended to Clayton Burger for polishing the context and the reviewers of this paper.

References

[BO07] Bogue, R.: Design for disassembly: a critical twenty-first century discipline. Assembly Automation, 27, 285-289, 2007.

[CD09] Crul, M.; Diehl, J.C.; Ryan, C.: Design for sustainability: a step-by-step approach. UNEP, Paris, 2009.

[GW15] Go, T.F.; Wahab, D.A.; Hishamuddin, H.: Multiple generation life-cycle for product sustainability: the way forward. Journal of Cleaner Production, 95, 16-29, 2015.

[IG10] Ilgin, M.A.; Gupta, S.M.: Environmentally conscious manufacturing and product recovery (ECMPRO): A review of the state of the art. Journal of Environmental Management, 91, 563-591, 2010.

[MV03] Maxwell, D.; Vorst, R.V.D.: Developing sustainable products and services. Journal of Cleaner Production, 11, 883-895, 2003.

[RI02] Ritchey, T.: Modelling complex socio-technical systems using morphological analysis. Adapted from an address to the Swedish Parliamentary IT Commission, Stockholm, 2002.

[RI06] Ritchey, T.: Modelling multi-hazard disaster reduction strategies with computer-aided morphological analysis. Proceedings of the 3rd International ISCRAM conference, Newark, NJ (USA), 2006.

[RI07] Ritchey, T.: Threat analysis for the transport of radioactive material. The 15th International Symposium on the Packaging and Transportation of Radioactive Materials, Miami, 2007.

[RR10] Ramani, K.; Ramanujan, D.; et al.: Integrated sustainable life cycle design: a review. Journal of Mechanical Design, 132, 1-15, 2010.

[RS02] Ritchey, T.; Stenstrom, M.; Eriksson, H.: Using morphological analysis for evaluating preparedness for accidents involving hazardous materials. The 4th International Conference for Local Authorities, Shanghai, 2002.

[ST13] Sakundarini, N.; Taha, Z.; Abdul-Rashid, S.H.; Ghazila, R.A.R.: Optimal multi-material selection for lightweight design of automotive body assembly incorporating recyclability. Materials and Design, 50, 846-857, 2013.

[SW15] Swedish Morphological Society (SWEMORPH).: Academic Research Support. http://www.swemorph.com/, 2003-2015.

[TC13] Tu. J.C.; Chiu, P.L.; Huang, Y.C.; Hsu, C.Y.: Influential factors and strategy of sustainable product development under corporate social responsibility in Taiwan. Mathematical Problems in Engineering, Article ID 303850, 2013.

[UF13] Umeda, Y.; Fukushige, S.; Mizuno, T.; Matsuyama, Y.: Generating design alternatives for increasing recyclability of products. CIRP Annals – Manufacturing Technology, 62, 135-138, 2013.

[ZW69] Zwicky, F.: Discovery, invention, research – through the morphological approach, Toronto: The Macmillan Company, 1969.

[ZX12] Zhang, X.M.; Xu, X.Y.; He. P.: New product desgin strategies with subsidy policies. Journal of System Science and System Engineering, 21, 365-371, 2012.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Umweltinformatik zwischen Nachhaltigkeit und Wandel (UINW 2015)

Environmental Informatics between Sustainability and Change

3. Workshop: Umweltinformatik zwischen Nachhaltigkeit und Wandel (UINW 2015)

Stefan Naumann[1], Kristina Voigt[2]

Abstract: Die Bedeutung der Informatik in den Naturwissenschaften, den Lebenswissenschaften, den Umweltwissenschaften und für die Energiewende ist unbestritten. Umweltinformatik ist ein wichtiges Feld der angewandten Informatik. Dies wurde bereits in den achtziger Jahren erkannt und in der Gründung des Fachausschuss Umweltinformatik 1986 umgesetzt. Ziel des Workshops ist der Erfahrungsaustausch zur Analyse und Lösung von Umweltproblemen mit Hilfe aktueller Methoden der Informatik und der Kommunikationstechnologien.

Keywords: Umweltinformatik, Nachhaltigkeit, Energie, Fachausschuss Umweltinformatik

1 Programmkomitee

- Hans-Knut Arndt, Magdeburg
- Werner Geiger, Karlsruhe
- Eva Kern, Lüneburg
- Stefan Naumann, Birkenfeld (Chair)
- Werner Pillmann, Wien
- Kristina Voigt, München (Chair)
- Jochen Wittmann, Berlin
- Volker Wohlgemuth, Berlin

2 Beiträge

Es wurden 8 Beiträge eingereicht, die nach eingehender Begutachtung sämtlich als für den Workshop geeignet bewertet wurden. Die Themen reichen von Beiträgen aus dem Bereich der Energie und der Wirtschaft über Fragen der Nachhaltigkeit von Software bis hin zur ökologischen Bewertung der Durchführung einer Informatik-Konferenz.

[1] Hochschule Trier, Umwelt-Campus Birkenfeld, Institut für Softwaresysteme, Postfach 1380, D-55761 Birkenfeld, s.naumann@umwelt-campus.de
[2] Dr. Kristina Voigt, Helmholtz Zentrum München, Institute of Computational Biology, Ingolstädter Landstr. 1, 85764 Neuherberg, kvoigt@helmholtz-muenchen.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Grundkonzeption einer ‚Kritischen Theorie der Hybridität' und Implikationen für „nachhaltige Wissenschaften"

Malte Lärz[1]

Abstract: Die nachfolgende Arbeit skizziert eine ‚Kritische Theorie der Hybridität' ausgehend der ersten Generation der Kritischen Theorie, der Akteur-Netzwerk-Theorie nach Latour und Callon sowie konzeptionellen Überlegungen zur Netzwerkgesellschaft nach Castells. So gilt es eine symmetrische Betrachtung hybrider Krisenphänomene und zu Grunde liegender Ursachen in der Wissens- und Technikproduktion moderner Gesellschaften vorzunehmen, sowie – ausgehend der angestellten Überlegungen – zu begründen, weshalb „nachhaltige Wissenschaften" eine deutlich stärkere Subjektorientierung und philosophische Durchdringung der Einzeldisziplinen benötigen.

Keywords: Kritische Theorie, Akteur-Netzwerk-Theorie, Hybridität, Nachhaltigkeit

1 Anamnesis der Genese – ‚Kritische Theorie der Hybridität'

Marx und Freud können *„[…] als die Begründer einer dritten Art von Wissenschaft gelten, die sich mit Institutionen der Seele und der Sozietät befasst, die den vergesellschafteten Individuen als ‚natürliche' erscheinen und sich dadurch gegen Veränderung immunisieren."* (Dahmer 2013, S. 1) Eine Bezugnahme auf dieses Wissenschaftsverständnis bedeutet pseudonatürliche Konstruktionen der Moderne zu kritisieren und sie durch Rekonstruktion ihrer Entstehungsgeschichte zu defetischisieren. Um solch ein Vorhaben zu leisten, gilt es, an die prägnanteste Charakteristik des Freud'schen und Marx'schen Projektes zu erinnern: *„[…] Anamnesis der Genese […]"*[2] (Adorno 1965, zitiert nach Sohn-Rethel 1989, S. 223) - Die Enträtselung der Sphingen, die den Weg zu einer gesellschaftlichen Assoziation verhindern, in der die demokratische Verwaltung über Technik und Gesellschaft in Versöhnung mit der Natur an die Stelle der Herrschaft über Menschen und Natur durch Technik getreten ist (vgl. Dahmer 2013, S. 1). *„Darum ist es an der Zeit, sich der eigentlichen Intention der psychoanalytischen Aufklärung zu erinnern: obsolet gewordenen Institutionen der Lebens- und Kulturgeschichte den Naturschein abzustreifen, um ihre Revision zu ermöglichen."* (ebd., S. 2) Solch eine Revision durch Restitution einer Kritischen Theorie innerhalb der modernen Industriegesellschaften ist nur durch eine ‚Kritische Theorie der Hybridität' zu leisten. Einerseits sind Gesellschaft und Natur zunehmend mit Technostrukturen durchzogen, andererseits tritt Technik den vergesellschafteten Individuen als zweite Natur entgegen (vgl. Böhme et Manzai 2003, S. 10). *„Kritische Theorie kann deshalb […] heute nicht mehr nur kritische Theorie der Gesellschaft sein,*

[1] Universität Lüneburg, Auf der Rübekuhle 21, 21335 Lüneburg, malte.laerz@stud.leuphana.de
[2] *„Historischer Materialismus ist Anamnesis der Genese."*

sondern sie muss zugleich kritische Theorie der Technik und der Natur sein." (ebd.) Das Wesen solch einer Theorie muss hybrid sein und gegenwärtig die mit der Geschichte gewachsenen und mit der Zukunft wachsenden Verknüpfungen zwischen Mensch, Natur und Technik im Angesicht der Endlichkeit des Planeten kontexttranszendent denken. Um dieser Hybridität zu genügen, bedient sich die vorliegende Grundkonzeption mit der Akteur-Netzwerk-Theorie einer Perspektive, die die symmetrische Koevolution von Mensch, Technik und Natur als abhängige Variablen in den Mittelpunkt der Gesellschaftskonstitution rückt (vgl. Schulz-Schäfer 2000, S. 195). Ausgehend theoretischer Überlegungen der ANT gilt es in den folgenden Fragmenten, *„[…] Technik als Lebensform darzustellen und die mit der Entstehung und Verwendung von Technik verbundenen Pathologien aufzuspüren und das praktische Ziel einer menschenwürdigen Gesellschaft […] nicht aus den Augen zu verlieren."* (Schmid Noerr 2003, S. 66)

2 Der Ökonomische Überbau und die Technik

Castells unterstreicht in seiner Konzeption einer Netzwerkgesellschaft, dass die kapitalistische Wirtschaftsform mit der stetigen Verbreitung der Computertechnik einen zunehmend informationellen und globalen Charakter annimmt. Sie ist *„[…] von ihrer Fähigkeit abhängig […], auf effiziente Weise wissensbasierte Informationen hervorzubringen, zu verarbeiten und anzuwenden."* (Castells 2003, S. 83) In dieser ökonomischen Abhängigkeit zur Rationalisierung von Wissen erhält dieses in Form der operablen Information einen instrumentellen Charakter (vgl. Adorno et Horkheimer 1947, S. 10-13). Im Zuge der Formalisierung und Nutzbarmachung von Wissen geriet die Mathematik zur systematischen Figur des Wissens und erzeugt das *„[…] Schema der Berechenbarkeit der Welt […]"*. (ebd., S. 13) Die Beherrschung von Mensch und Natur durch Wissen und ihre realweltliche Manifestation in Form von Technik *„[…] zielt nicht auf Begriffe und Bilder, nicht auf das Glück der Einsicht, sondern auf Methode, Ausnutzung der Arbeit anderer, Kapital."* (ebd., S. 10) Ausgehend ihrer instrumentellen Form existiert Technik ebenso immateriell als Sozio- oder Psychotechnik, die darauf abzielt, *„[…] gesellschaftliche Prozesse wie Naturprozesse verfügbar […]"* (Habermas 1963, S. 481) zu machen, dementsprechend menschliche Handlungen durch Empirie ökonomisch zu erschließen (vgl. Adorno et Horkheimer 1947, S. 172) und schlussendlich beherrschbar zu machen (ebd., S. 13).

Die Einbindung der Technik und des Wissens in das Wirtschaftssystem macht sie ebenso demokratisch wie die Ökonomie, die sie hervorbringt (vgl. ebd., S. 10). Da ihre Verfügbarkeit über Kapital geregelt wird und ihre Produktion zweckorientiert geschieht, kennen Technik und Wissen keine Grenzen, *[…] weder in der Versklavung der Kreatur noch in der Willfährigkeit gegen die Herren der Welt […]* (ebd., S.10). Im Zuge der Globalisierung, dem Entstehen eines globalen Systems der internationalen Arbeitsteilung und Kommunikation von miteinander *„[…] verknüpften Segmenten von Volkswirtschaften […]"* (Castells 2003, S. 83), somit einer immer komplexer werdenden Logik des Netzwerkes aus Ökonomie, Wissen und Technik, entsteht eine *„[…] zunehmende Distanz zwischen Globalisierung und Identität […]."* (ebd., S.24) Der

Wandel des globalen ökonomischen Überbaus vollzieht sich durch rekursive Anwendung von Wissen und Technik auf Selbige (vgl. ebd., S.16, S.106) und führt zu einer exponentiellen Beschleunigung des technologischen Fortschritts (vgl. ebd. 17ff.), ökonomischem Wachstum und zu zunehmender *„[…] Konvergenz spezifischer Technologien zu einem hochgradig integrierten System […]"* (ebd, S. 77), welches zur Verknüpfung und Absorption älterer, separierter Technologiestränge zu einem einheitlichen Strang drängt. (vgl. ebd.) *Produktivität und Wachstumspotential dieses Systems stabilisieren die Gesellschaft und halten den technologischen Fortschritt im Rahmen von Herrschaft."* (Marcuse 1967, S. 19) Mit dem durch Konvergenz, Technik und Wissen erzeugten Wachstum des ökonomischen Netzwerkes vergrößern sich die Kosten, nicht zu partizipieren exponentiell (vgl. Castells 2003, S. 76), *„[…] weil die Chancen sich vermindern, andere außerhalb des Netzwerks zu erreichen."* (ebd.) Ein Austritt aus diesem Netzwerk der rekursiven Technik- und Wissensproduktion führt zum Ausschluss von den Ressourcen *„[…] Kapital, Information, Technologie, Güter, Dienstleistungen, qualifizierte Arbeit […]"* (ebd., S. 156f.)

3 Technik als Lebensform: Instrumentelle Moralität

Ausgehend des Wachstums des globalen ökonomischen Netzwerks wird die Herrschaft über Mensch und Natur durch die Produktion von Technik und Wissen weiter vorangetrieben. Wissens- und Technikentwicklung geschieht hierbei als Resultat der Verknüpfung der heterogenen Komponenten zu (Sub-)Netzwerken. *„Die Eigenschaften und Verhaltensweisen der beteiligten belebten und unbelebten Natur, die der involvierten technischen Artefakte und die der betreffenden sozialen Akteure […] – sie alle sind Gegenstand und Resultat der wechselseitigen Relationierung im Netzwerk. Und zugleich werden sie allesamt als Handlungssubjekte solcher Prozesse betrachtet."* (Schulz-Schäffer 2000: S. 188) Im Gegensatz zu den belebten Aktanten, sind technische Artefakte in ihren Wesenszügen und Entitäten durch das Netzwerk schwerer veränderbar (vgl. Latour 1996, S. 55-60) und somit ausgehend ihrer funktionalen Moralität präskriptiv wirkend. *„Präskription ist die ethische und moralische Dimension der Mechanismen. Auch wenn die Moralisten dies beständig beklagen, ist kein Mensch so unnachgiebig moralisch wie eine Maschine […]. Mit der Ansiedlung von Nicht-Menschen bleibt die Summe der Moralität nicht gleich, sondern wächst beträchtlich. Hier sind sie, die versteckten und verachteten sozialen Massen, aus denen unsere Moralität besteht."* (Latour 1992, S. 232) Die Moralität der Technik erzeugt eine besondere Mitwirkung an der Entstehung, Aufrechterhaltung und Entwicklung gesellschaftlicher Assoziationen: *„Ohne die Partizipation […] von Nicht-Menschen, insbesondere von Maschinen und Artefakten, ist kein soziales Leben denkbar. Ohne sie würden wir wie Paviane leben."* (Callon et Latour 1992, S. 359) So übernimmt Technik ausgehend ihrer Eigenschaften in Gesellschaften die Funktion einer ‚Black Box', die - reproduziert aus historischen Assoziationen - Arbeit und Wissen unzähliger weiterer Aktantennetzwerke enthält. Bei Öffnung einer ‚Black Box' erscheinen unzählige weitere Objekte, die ihrerseits wiederum assoziativ zerlegbar sind. Existierende Technik als zeitweilig stabiler Zustand eines Netzwerkes schüttelt in dieser Komplexität für

gewöhnlich ihre Entstehungsgeschichte ab und macht Identitäten und Beziehungsmuster der Aktanten selbstevident (vgl. Callon 1991, S. 145, vgl. Schulz-Schäfer 2000, S. 204): Die meisten Handlungen im Umgang mit Technik werden zu *„[…] stillschweigenden, gewohnten, (in menschliche und nicht-menschliche) Körper eingebauten Handlungen […]"* (Latour 1988, S. 308). Technik geriet zur stabilen Konvention, die ausgehend ihrer Beschaffenheit vorgibt, wie sie zu nutzen ist – *„[…] im Wesentlichen ist die Macht der Maschine nur die aufgespeicherte und projektierte Macht […]"* (Marcuse 1967, S. 23) früherer gesellschaftlicher Assoziationen.

Ausgehend des ökonomischen Überbaus sowie der rekursiven Wissens- und Technikproduktion stehen Mensch und Maschine im Prozess der Arbeitsteilung in direkter Konkurrenz zueinander, wobei die Maschinen ausgehend des technologischen Fortschritts und ihrer stärkeren Moralität dem Menschen zunehmend überlegen sind und somit Arbeitslogik und -modalität definieren: *„Das rohe Faktum, dass die […] Gewalt der Maschine die des Individuums […] übertrifft, macht die Maschine in jeder Gesellschaft, deren grundlegende Organisation, die des maschinellen Prozesses ist, zum wirksamsten Instrument."* (ebd., S. 23) Die Arbeitsteilung, der ökonomisch induzierte Konkurrenzkampf zwischen belebten und unbelebten Aktanten, drängt den Menschen zur Selbstbehauptung durch Anpassung an seine zweite Natur, die ihm zunehmend natürlich erscheint. *„Je weiter aber der Prozess der Selbsterhaltung durch bürgerliche Arbeitsteilung geleistet wird, umso mehr erzwingt er die Selbstentäußerung der Individuen, die sich an Leib und Seele nach der technischen Apparatur zu formen haben."* (Adorno et Horkheimer 1947, S. 37) Die formalisierte Logik der sich durch Technik- und Wissensentwicklung evolutionierenden Maschinen wird zum Programm des instrumentellen Denkens: *„Denken verdinglicht sich zu einem selbsttätig ablaufenden Prozess der Maschine nacheifernd […]*. (ebd., S. 31) Die im Prozess der arbeitsteiligen Technik- und Wissensproduktion assoziierten belebten und unbelebten Entitäten mögen ‚natürlich' zwar auseinanderweisen (vgl. Latour 2006, S. 392), die Stabilität der Technik jedoch besitzt die Möglichkeit, die gebildeten Assoziationen und die Gesellschaft *„[…] als beständiges Ganzes zusammenzuhalten."* (ebd.) Das Soziale und Kulturelle transformiert sich in eine Zirkulation (vgl. ebd., S. 565), die ausgehend der rekursiven Technik- und Wissensproduktion katalysiert wird. Technik erhält im Zuge der voranschreitenden Ökonomisierung der Welt ihre instrumentelle Präskription. Ausgehend ihres instrumentellen Charakters tendiert sie, je weiter das Wachstum der Ökonomie voranschreitet, die gesellschaftliche Moralität ins dialektisch Negative: *„[…] Konkretisierungen der Moral […] tragen repressive Züge."* (Adorno 1966, S. 253)

4 Die Computertechnik als universeller Technologiestrang

„Angesichts der totalitären Züge dieser Gesellschaft lässt sich der traditionelle Begriff der Neutralität der Technik nicht mehr aufrechterhalten. Technik als solche kann nicht vom Gebrauch abgelöst werden […]; die technische Gesellschaft ist ein Herrschaftssystem, das bereits im Begriff und Aufbau der Techniken ist." (Marcuse 1967, S. 18) Die Computertechnik stellt den Technologiestrang dar, der durch seine

Absorptionsfähigkeit immaterielle und materielle Techniken miteinander verknüpft und die globale, zeitgenaue Kommunikation und Verwaltung der Ökonomie ermöglicht. *„Der Computer […] ist ein Instrument, das in den Dienst gezwungen wurde, um die […] reaktionärsten ideologischen Strömungen des gegenwärtigen Zeitalters zu rationalisieren, zu unterstützen und am Leben zu halten. […] Lange bevor Computer zum Fetisch wurden und der Verfinsterung der Vernunft eine konkrete Gestalt gaben, hat Horkheimer uns die notwendige Perspektive eröffnet."* (Weizenbaum 1976, S. 327) Ihre besonderen Eigenschaften erhält die digitale Infrastruktur durch die binäre Sprache. So notierte Zuse bereits am *„[…] 19. Juni 1937: Erkenntnis, dass es Elementaroperationen gibt, in die sich sämtliche Rechen- und Denkoperationen auflösen lassen."* (Zuse 1937, zitiert nach Heintz 1993, S. 211) In dieser Auflösung ins mathematisch Darstellbare erhält das Binäre im arbeitsteiligen Prozess der Wissens- und Technikproduktion eine universelle Form, die durch die Ökonomie im Gewand des Fortschritts zunehmend instrumentalisiert wird. *„[…] Ideal ist das System, aus dem alles und jedes folgt. […] Die bürgerliche Gesellschaft ist beherrscht vom Äquivalent. Sie macht Ungleichnamiges komparabel, indem sie es auf abstrakte Größen reduziert."* (Adorno et Horkheimer 1947, S.13) Im Zuge dieser Reduktion, die Interaktion im ökonomischen System in vermeintlich beschreibbare Qualitäten und Quantitäten umwandelt, *„[…] bringt der technische Fortschritt Lebensformen (und solche Macht) hervor, welche […] allen Protest im Namen […] der Freiheit zu besiegen oder zu wiederlegen scheinen."* (Marcuse 1967, S. 14) Die subtile Form der technischen Herrschaft findet sich im Wesen der Empirie wieder und verfeinert sich ausgehend der rekursiven Technik- und Wissensproduktion bei gleichzeitiger Ausweitung des empirischen Systems der Aktanten. *„Und just weil es vollständig durch seine empirische Erscheinung definiert ist, kann das Ding analytisch in unterschiedliche Qualitäten und Quantitäten zerlegt und von einem technischen System absorbiert werden, das es ganz anderen Zwecken unterwirft […]. Menschliche Wesen sind ins System einbezogen, sie sind auch nur ein Element in dem fungiblen Stoff, der den Sozialapparat ausmacht."* (Feenberg 2003, S. 47, vgl. Marcuse 1967, S. 163-168)

Die sich ausbreitende technische Infrastruktur und ihre Einbindung in die Ökonomie erhöht das ‚Enrollment' der Aktanten - *„[…] die Art und Weise, in der ein Satz zueinander stehender Rollen definiert und den Akteuren zugeschrieben wird, die sie akzeptieren."* (Callon 1986, S. 211) Die instrumentellen Verbindungen des ökonomischen Überbaus *„[…] halten nur dann zusammen, wenn die verschiedenen betroffenen Entitäten […] die ihnen zugeschriebene Rolle akzeptieren […]"* (Callon 1987, S. 93), dementsprechend *„[…] den erforderlichen (Re-)Definitionen entweder keinen Widerstand entgegensetzen oder dazu gebracht werden können, ihren Widerstand aufzugeben, oder aber überhaupt außer Stande gesetzt werden, sich ihnen zu entziehen."* (Schulz-Schäfer 2000, S. 201) Die technische Infrastruktur der Ökonomie gerät zur größten und stabilsten Assoziation, die ausgehend ihres Wachstumszwangs totale ‚Konvergenz' und ‚Irreversibilität' der interagierenden Aktanten anstrebt. Die durch Technik und vergesellschaftete Individuen zweckorientiert zugerichtete Welt tendiert zum geschlossenen System einer sich durch Technik verwaltenden Welt, welche *„[…] die Unmöglichkeit der Rückkehr zu konkurrierenden Übersetzungen […]"* (Callon 1991,

S. 151) anstrebt sowie deren Konvergenz bedingt, dass, je mehr Aktanten zusammenarbeiten, umso weniger zweifelhaft deren Status ist. (vgl. ebd., S. 148) Die Irreversibilität dieser Entwicklung *„[…] nimmt in dem Maße zu, in dem jedes Element […] in ein Bündel wechselseitiger Beziehungen eingeschrieben ist […]“* (ebd., S. 150), also in dem Maße, in dem die vernetzte Technisierung voranschreitet. Callon führt als Beispiel solch eines Netzwerkes den Turm zu Babel an - aus Perspektive einer kritischen Theorie die mythologische Überhebung des Menschen über die zu entzaubernde Natur durch die instrumentelle Vernunft: *„Auf, formen wir Lehmziegel und brennen wir sie zu Backsteinen. So dienten ihnen gebrannte Ziegel als Steine und Erdpech als Mörtel. Dann sagten sie: Auf, bauen wir uns eine Stadt und einen Turm mit einer Spitze bis zum Himmel und machen wir uns damit einen Namen […]. Da stieg der Herr herab, um sich Stadt und Turm anzusehen, die die Menschenkinder bauten. Er sprach: Seht nur, ein Volk sind sie und eine Sprache haben sie alle. Und das ist erst der Anfang ihres Tuns. Jetzt wird ihnen nichts mehr unerreichbar sein […].“* (Genesis 11, 3-7)

5 Hybridität als globales Netzwerk: Turmbau zu Babel

Noch immer steht der Turmbau zu Babel für die menschliche Hybris, sich mithilfe der Technik über die Gesetze der Natur erheben zu wollen. Die ‚Ziegel‘ werden mittlerweile durch den Computer, das ‚Erdpech‘ als sozialer Kitt durch die Vernetzung und die ‚gemeinsame Sprache‘ durch die Übersetzungen in das Binärsystem abgelöst: Die gesellschaftliche Assoziation ‚Turmbau‘ wächst und mit ihr die Tendenzen der Welt zum geschlossenen, arbeitsteiligen System. *„Die Arbeitsteilung zu der sich Herrschaft gesellschaftlich entfaltet, dient dem beherrschten Ganzen zur Selbsterhaltung. […] Die Herrschaft tritt dem Einzelnen als das Allgemeine gegenüber, als die Vernunft in der Wirklichkeit.“* (Adorno et Horkheimer 1947, S. 31) Während der ökonomische Überbau die Technik- und Wissensentwicklung weiter beschleunigt, hybridisiert der voranschreitende Prozess der Mensch- und Naturaneignung in Form des Fortschritts die Erde. So werden zwar in Folge dessen *[…] Natur- und Gesellschaftsprozesse […] wissenschaftlich und technisch planbarer […]*, jedoch zugleich auch *[…] tendenziell unvorhersehbar und unbeherrschbar.“* (Becker et Jahn 1989, S. 18) Diese *[…] ‚Krise der Mittel‘ im Sinne von Überkomplexität und Kontingenz […]* (ebd.) überlagert die Zweckrationalität des Wissens und der Technik und verleiht ihr einen krisenhaften Charakter, der sich in hybriden Krisenphänomenen *[…] auf das individuelle und gesellschaftliche Alltagsbewusstsein […]“* (ebd., S.19) niederschlägt. *„Das Ozonloch ist zu sozial […], um wirklich Natur zu sein, die Strategie von Firmen und Staatschefs zu sehr angewiesen auf chemische Reaktionen, um allein auf Macht und Interesse reduziert werden zu können, der Diskurs der Ökosphäre zu real und zu sozial, um ganz in Bedeutungseffekten aufzugehen.“* (Latour 2008, S. 14) Das Spezifikum der Moderne ist die konsequente Fehlwahrnehmung und die unkontrollierte Vermehrung dieser Hybride durch ‚im Labor‘ geschaffene Technik (vgl. ebd., S. 49).

Aufgrund der gestiegenen Komplexität der Arbeitsteilung, der wachsenden Vernetzung und der sich offenbarenden globalen Wirkzusammenhänge der Umweltmedien existieren

in Folge der technischen Entwicklung *„[…] hybride Krisenphänomene, in denen gesellschaftliche und natürliche Prozesse sich überlagern, interferieren und eine komplexe Krisendynamik entwickeln."* (Becker et Jahn 2006, S. 169) Ausgehend dieser Dynamiken, die sich bspw. im Klimawandel, der Mikroplastik- oder der Peak-Everything-Problematik niederschlagen, wird die Stabilität der Ökonomie zunehmend bedroht. Die gegen unendlich strebende wissens- und technikbasierende Ausdifferenzierung und Produktion der wachstumsgetriebenen Gesellschaften lässt Enzensbergers global-ökologische Hypothese, die den Zusammenbruch der Biosphäre und den Wegfall der benötigten Ressourcen (vgl. Enzensberger 1973, S. 11f.), zunehmend wahrscheinlich erscheinen. Ausgehend ihrer Anerkennung zwingt die global-ökologische Hypothese zur Neubewertung der hybriden Krisenerscheinungen, da die *„[…] industrialisierten Gesellschaften der Erde ökologische Widersprüche produzieren, die in absehbarer Zeit zum Zusammenbruch führen."* (ebd., S. 12) Der Turmbau ist, so denn die Gesetze der Natur durch Technik nicht gebrochen werden, zwangsläufig zum Scheitern verurteilt. *„Dient nicht die Bedrohung durch eine […] Katastrophe […] gerade diejenigen Kräfte zu schützen, die diese Gefahr verewigen? Die Anstrengungen, eine solche Katastrophe zu verhindern, überschatten die Suche nach ihren etwaigen Ursachen in der gegenwärtigen Industriegesellschaft."* (Marcuse 1967, S. 11) Während die externe Grenze des Kapitals in Form der endlich belastbaren Naturressourcen weiterhin unverrückbar existiert, wird die Machbarkeit des unendlichen Wachstums […] *mit den problemlösenden Wirkungen eines technischen Fortschritts begründet, der noch gar nicht eingetreten ist. Und es ist nicht beweisbar, dass er jemals eintreten wird."* (Paech 2011, S. 4) Instrumente wie die Klimasimulation, die einst geschaffen, die global-ökologische Hypothese zu untermauern (vgl. Becker et Jahn 1989, S. 16), erhalten im Zuge der zweckorientierten Rezeption angesichts der Überkomplexität einen quasi-religiösen Charakter. Die praktische Regulierung der Krisenphänomene durch Konsumentenempfehlungen, 2-Grad-Zielen, Monitorings etc. bei anhaltender rekursiver Wissens- und Technikproduktion bestärken den Mythos von der Berechenbarkeit der Welt und liefern die benötigten Opiate für die vergesellschafteten Individuen im voranschreitenden Prozess der Hybridisierung. *„Als Begleitmusik erklingt die Hymne einer demnächst alle Probleme lösenden Innovationswelle. Diese […] Fortschrittsvisionen […] erfüllen eine ausschließlich kommunikative Funktion: Sie verleihen einem auf Maßlosigkeit gründenden Wohlstandsmodell eine pseudo-wissenschaftliche Legitimation.* (Paech 2011, S. 4)

6 Datenerhebung zur Stabilisierung der hybriden Welt

Während der Turmbau angesichts der hybriden Krisenphänomene zunehmend ins Wanken gerät, werkeln unzählige Wissenschaftler im Kontext des Green New Deals am Versuch die unweigerliche Grenzapproximation bei stetiger Kapitalakkumulation hinauszuzögern (vgl. Mahnkopf 2013, S. 8f.). Ihre Bemühungen in dieser Sysiphos-Aufgabe gehorchen nicht dem Gegenstand der Krise, sondern dem Primat der Kapitalakkumulation. Angesichts der global-ökologischen Hypothese wandelt sich die empirische Erschließung des Planeten anhand der ökonomischen und

naturwissenschaftlichen Indikatorenkombination zunehmend zum Maßstab einer verwalteten Welt. *„Jede Ökobilanzierung ist mit Schwierigkeiten der Informationsbeschaffung sowie Abgrenzung, Zurechnung und Gewichtung einzelner Wirkungen behaftet. Aber sie ist alternativlos.“* (Paech 2011, S. 5) Diese Alternativlosigkeit wird deutlicher, je weiter die Verwüstungen des Planeten voranschreiten und je dünner das Drahtseil, das die ökonomische und naturwissenschaftliche Indikatorenkombination bei zunehmender hybrider Komplexität und Krisenhaftigkeit spannt, wird. Während die Nutzbarmachung der Information im technischen System im Angesicht der global-ökologischen Hypothese alternativlos erscheint, erzeugt die digitale Technik ausgehend ihrer formalisierenden Arbeitsweise die Angebotsseite und ermöglicht den Tauschhandel zwischen vergesellschafteten Individuen und Technik. Das Internet wandelt sich zunehmend zum Marktplatz der Informationsbeschaffung und ausgehend der wachsenden Sensorik, der mathematischen Formalisierung und der zunehmenden Hybridisierung in den Raum der Informationsströme jeglicher Art. Die Vermessung der Erde richtet sich schlussendlich gegen den Menschen: *„Wo die Entwicklung der Maschine in die der Herrschaftsmaschinerie schon umgeschlagen ist, so dass technische und gesellschaftliche Tendenz […] in der totalen Erfassung der Menschen konvergieren […],* (Adorno et Horkheimer 1947, S. 43) nimmt die Kontrolle über die Menschen und ihre Bedürfnisse überhand und macht sie zu Objekten totalitärer Beherrschung durch den Produktionsapparat (vgl. Marcuse 1967, S. 17).

7 Subjektorientierung im Sinne

Durch die wachstumsgetriebene Ökonomie und der einhergehenden Technisierung erscheint der Weg zu nachhaltigen Gesellschaftsentwürfen verstellt, jedoch umso zwingend notwendiger, je weiter die Hybridisierung der Welt durch Technik vorrangeschritten ist. Dieses Paradox birgt eine gefährliche Dialektik in sich, die im Angesicht sinkender ökologisch-sozialer Systemstabilitäten zu technischen Lösungen drängt, die menschenrechtliche Errungenschaften der modernen Gesellschaften zunehmend einschränken. *„Das Verzweifelte, daß die Praxis, auf die es ankäme, verstellt ist, gewährt paradox die Atempause zum Denken, die nicht zu nutzen praktischer Frevel wäre.” (Adorno 1966, S. 243)* Aus Perspektive der skizzierten Überlegungen einer ‚Kritischen Theorie der Hybridität gilt es „nachhaltigen Wissenschaften“ zu einer stärkeren Subjektorientierung zu raten. So gilt es eine *“[...] eine Diktatur des planvollen Arbeitens über das Nebeneinander von philosophischer Konstruktion und Empirie in der Gesellschaftslehre zu errichten […]”* (Horkheimer 1931, S. 12), die sich kritisch mit dem Verhältnissen von Technik, Mensch und Natur auseinandersetzt. Da innerhalb dieser Verhältnisse Technik eine spezielle Funktion übernimmt, gilt es *„[…] einen positiven Begriff von ihr vorzubereiten, der sie aus ihrer Verstrickung in blinder Herrschaft löst […].“* (Adorno et Horkheimer 1947, S. 21)

Literaturverzeichnis

[AH47] Adorno, T.W., Horkheimer, M.: Dialektik der Aufklärung – Philosophische Fragmente. 15.te Auflage (2004), Fischer Taschenbuchverlag, Frankfurt am Main, 1947.

[Ad66] Adorno, T.W.: Negative Dialektik - Jargon der Eigentlichkeit. 7. Aufl. (2012), Suhrkamp Verlag, Frankfurt am Main, 1966.

[BJ89] Becker, E.; Jahn, T.: Soziale Ökologie als Krisenwissenschaft – Sozial-ökologische Arbeitspapiere 1. IKO-Verlag, Frankfurt am Main, 1989.

[BJ06] Becker, E.; Jahn, T.: Soziale Ökologie. Grundzüge einer Wissenschaft von den gesellschaftlichen Naturverhältnissen. Campus Verlag, Frankfurt am Main, New York, 2006.

[BM03] Böhme, G.; Manzai, A.: Eine kritische Theorie der Technik und der Natur. Willhelm Fink Verlag, München, 2003.

[Ca03] Castells, M.: Das Informationszeitalter. Bd.1: Der Aufstieg der Netzwerkgesellschaft. Leske+Budrich, Opladen, 2003.

[Ca86] Callon, M.: Some Elements of a Sociology of Translation. Domestication of the Scallops and the Fishermen of St. Brieuc Bay. In: Power, Actions and Belief: A new Sociology of Knowledge. Routledge & Kegan Paul, London, S. 196-233, 1986.

[Ca87] Callon, M.: Society in the Making: The Study of Technology as a Tool for Sociological Analysis. In: The Social Construction of Technological Systems: New Directions in the Sociology and and History of Technolgy. MIT Press, London, S. 83-103, 1987.

[Ca91] Callon, M.: Techno-economic networks and irreversibility. In: A Sociology of Monsters: Essays on Power, Technology and Domination. Routledge & Kegan Paul, London, S. 132-165, 1991.

[CL92] Callon, M.; Latour, B.: Don't throw the baby out with the bath school! A reply to Collins and Yearly. In: Pickering, A. (Hrsg.): Science as Practices and Culture. Chicago, University of Chicago Press, S. 343-368, 1992.

[Da13] Dahmer, H.: Restitution einer "Kritischen Theorie". In: Kritiknetz - Zeitschrift für Kritische Theorie der Gesellschaft, 2013.

[En73] Enzensberger, H.M.: Zur Kritik der politischen Ökologie. In: Ökologie und Politik oder die Zukunft der Industrialisierung (Kursbuch 33). Berlin, 1973.

[Fe03] Feenberg, A.: Heidegger und Marcuse: Zerfall und Rettung der Aufklärung. In: Böhme, G.; Manzai, A. (Hrsg.): Eine kritische Theorie der Technik und der Natur, Willhelm Fink Verlag, München, S. 39-53, 2003.

[Ha68] Habermas, J..: Analytische Wissenschaftstheorie und Dialektik – Ein Nachtrag zur Kontroverse zwischen Popper und Adorno. In: Horkheimer, M. (hrsg.): Zeugnisse. Festschrift für Theodor W. Adorno. Europäische Verlagsanstalt, Frankfurt am Main, 1968.

[He93] Die Herrschaft der Regel. Zur Grundlagengeschichte des Computers. Suhrkamp Verlag, Frankfurt am Main, 1993.

[Ho31] Die gegenwärtige Lage der Sozialphilosophie und die Aufgaben eines Instituts für Sozialforschung. In: Horkheimer, M.: Sozialphilosophische Studien. Aufsätze, Reden und Vorträge 1930-1972. 2. Aufl. (1981), Fischer Taschenbuch Verlag, Frankfurt am Main, 1931.

[La88] Latour, B.: Mixing Humans and Nonhumans together: The Sociology of a door-closer. In: Social Problems, Vol. 35, No.3, London, S. 298-310.

[La92] Latour, B.: Where are the missing masses? The sociology of a few mundane artifacts. In: Bijker, W., Law, J. (Hrsg.): Shaping Technology / Building Society: Studies in Sociotechnical Change, S. 225-258, 1992.

[La96] Latour, B.: Der Berliner Schlüssel. Erkundungen eines Liebhabers der Wissenschaft. Berlin, 1996.

[La06] Latour, B.: ANT, in: Belliger, A.; Krieger, D.J. (Hrsg.): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Transcript-Verlag, Bielefeld, 2006.

[La08] Latour, B.: Wir sind nie modern gewesen – Versuch einer symmetrischen Anthropologie. Suhrkamp Verlag, Frankfurt am Main, 2008.

[Ma67] Marcuse, H.: Der eindimensionale Mensch. Studien zur Ideologie der fortgeschrittenen Industriegesellschaft. Luchterhand, Neuwied, 1967.

[Ma13] Mahnkopf, B.: Peak Everything – Peak Capitalism – Folgen der sozial-ökologischen Krise für die Dynamik des historischen Kapitalismus. Working Paper der DFG-KollegforscherInnengruppe Postwachstumsgesellschaften, Nr. 02/2013, Jena, 2013.

[Pa11] Paech, N. (2011): Adios, Konsumwohlstand: Vom Desaster der Nachhaltigkeitskommunikation und den Möglichkeiten der Suffizienz. In: Science Garden – Magazin für junge Forschung, S. 1-20.

[Sc00] Schulz-Schäfer, I.: Akteurs-Netzwerk-Theorie. Zur Koevolution von Gesellschaft, Natur und Technik, In: Schulz-Schäfer, I.: Soziale Netze, Konzepte und Methoden der sozialwissenschaftlichen Netzwerkforschung. München, S. 187-211, 2000.

[Sc03] Schmid Noerr, G.: Technik und Technikkritik im Denken Max Horkheimers, in: Böhme, G.; Manzai, A. (Hrsg.): Eine kritische Theorie der Technik und der Natur, Willhelm Fink Verlag, München, S. 55-67, 2003.

[So89] Sohn-Rethel, A.: Geistige und körperliche Arbeit. Weinheim, 1989.

[We76] Weizenbaum, J.: Die Macht der Computer und die Ohnmacht der Vernunft. Suhrkamp Verlag, Frankfurt am Main, 1976.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

ON THE ENVIRONMENTAL FOOTPRINT OF AN IS CONFERENCE

Florian Stiel[1], Frank Teuteberg[2]

Abstract: The IS research community has paid significant attention to climate change and other ecological problems. However, the communities own environmental footprint has been subject to little research so far. This work reports on the results of a Life Cycle Assessment that was conducted to identify the main determinants for the environmental footprint of an IS conference. It brings up the painful subject that also the scientific community has its substantial contribution to environmental degradation. Our work suggests that this contribution can be significantly reduced by only a few very effective instruments. Scientists are encouraged to enter the discourse on the environmental footprint of academic work in order to enhance Green IS research.

Keywords: Green IS, Environmental Footprint, Life Cycle Assessment, Academic Practice, Greenhouse Gases, Water Depletion, Emission of Toxic Substances, Umberto Software

1 Introduction

The impact of information systems and information technology on our modern society and organizations has been a controversial topic for many decades [OH02, We72]. Sustainable development that "meets the needs of the present without compromising the ability of future generations to meet their own needs" [Wc87] is still a grand challenge of our time that information systems research is called to accept [Br13, HC12]. Our work was inspired by [Br13], particularly by the directive to set examples for sustainable academic practice. They suggest that " [IS] scholars need to be more aware of the environmental footprint of their work"[Br13]. Recent studies suggest that attending conferences is one of the main sources of academic environmental impact [HH02, SL13, St08], such as greenhouse gas (GHG) emissions. We take a closer look at the environmental footprint of an IS conference in order to inform IS scholars about the environmental impact of their work behavior by answering the research question:

1. *What are the main determinants of the environmental impact of an IS conference?*

To answer this research question the *life cycle assessment* (LCA) method [Is09a, Is09b] was chosen. This work makes use of predictive and explorative scenarios [Bö06] in order to rate the environmental impact of future IS conferences [Fi09]. A special focus is put on dematerialization, a potential instrument used to reduce a conferences

[1] Universität Osnabrück, Fachgebiet Unternehmensrechnung und Wirtschaftsinformatik, Katharinenstr. 1, 49074 Osnabrück, florian.stiel@uni-osnabrueck.de
[2] Universität Osnabrück, Fachgebiet Unternehmensrechnung und Wirtschaftsinformatik, Katharinenstr. 1, 49074 Osnabrück, frank.teuteberg@uni-osnabrueck.de

environmental impact [Br11]. A distinct explorative scenario is analyzed in order to answer the following research question:

2. *What are effective instruments to reduce the environmental impact of a conference?*

The paper is structured as follows: The results of recent work are reviewed in section 2. Section 3 presents the method of this work. The results of this work are then presented and discussed in section 4. Finally, section 5 encompasses the concluding remarks as well as prospects of future work in the field.

2 Recent Work

Two literature searches were conducted by using the guidelines of [Co88], [WW02] and [Br09] in order to identify recent environmental footprints. The analysis is documented by using concept matrixes [WW02]. The search strings (Work OR Conference OR "Research Travel*") AND ("Carbon Footprint" OR "Life Cycle Assessment" OR "Environment* Impact") and (Paper OR Dematerialization OR "Digital Media") AND ("Life Cycle Assessment" OR "Carbon Footprint") were used querying the databases ScienceDirect, ISI webofscience CoreCollecion and AIS electronic library. Moreover, a backward search has been performed in order to identify additional sources. The first search revealed 6 results as shown in Tab. 1. A similar matrix was compiled for the second literature search. The results are depicted in Tab. 2.

Author(s), Year of Publication	Reference value	Metric	Subject area[3]	Scenario Analysis
[HH02]	1 conference	EIP[4]	Environmental Science	Yes
[MC07]	1 meeting	GHG	Medicine	Yes
[St08]	1 researcher/year	GHG	Earth and Planetary Sciences	No
[CHB12]	1 conference	GHG	Social Science/ Computer Science	Yes
[AAM13]	1 PhD project	GHG	Environmental Science	Yes
[SL13]	1 conference trip	GHG	Biochemistry/ Medicine	Yes

Tab. 1: Recent work on the environmental impact of science

Author, Year of Publication	Reference value	Metric	Subject area	Scenario Analysis
[GK02]	1 article	Energy	Environmental Science	Yes
[Mo10]	1 newspaper	Full LCA	Environmental Science	Yes
[TP10]	1 invoice	GHG	Information Systems	Yes
[MBF11]	1 book	Full LCA	Environmental Science	Yes
[MRS13]	1 managed form	Full LCA	Environmental Science	Yes
[ZZ13]	28 600 books	GHG, Energy	Environmental Science	Yes

Tab. 2: Recent work on the environmental impact of dematerialization and digital media

[3] Subject areas according to http://www.scimagojr.com
[4] Eco-indicator points [Mi00]

The literature documents an interest for the environmental impact of science in various subject areas, particularly during the last 2 years. Especially GHG emissions are of major interest. [HH02] suggest that the focus on GHG is reasonable since the choice of metric in their opinion does not significantly affect the decision-relevant aspects of the results. Moreover, all authors agree that traveling is the main source of science related environmental damage. [St08] and [CHB12] therefore introduce detailed mathematical models for the estimation of CO_2 emissions by travelling distance. A contrasting picture is drawn by the research on the environmental impact of dematerialization (paper print vs. digital media). The results of [MBF11] show that the choice of different metrics can turn the results of an environmental assessment into the direct opposite: Comparing e-books and paper books by the associated GHG emission leads to the assumption that the use of an e-books is beneficial when around 25 paper books are substituted. Using freshwater ecotoxicity as the leading metric the environmental trade-off for using paper books is reached at around 400 books. This contrasting use of metrics is of particular interest concerning RQ 2. Also the results of [MBF11], [Mo10] and [MRS13] suggest that a focus on GHG emissions is not sufficient to adequately support decision making. Moreover, the literature exposes that it is doubtful whether mathematical models for the calculation of traveling related GHG emissions are sufficient in order to answer RQ 1. The literature shows that the knowledge base concerning the environmental footprint of science is interspersed throughout different subject areas. Nevertheless, data, models as well as premises, simplifications and allocations from recent work provided a promising starting position for this research.

3 Methodology and Data

The goal and scope of this work was to explore environmental footprint associated with a scientific conference and to suggest effective instruments to establish sustainable academic practices within the IS community. The inventory analysis includes the collection of data, the model implementation, model parameterization as well as parameter variation (scenarios) [Gu02, Is09a]. The analysis involves a computer based calculation using the LCA software tool Umberto[5], based on petri-nets [Mu89], a tool that is frequently applied for LCA studies. Within the calculation we used data, models and instruments from two different sources: 1. specific data and specific partial models from recent work (c.f. section 2), 2. generic data for downstream activities from the leading LCA database ecoinvent[6]. Due to the high number of influencing variables, we made use of object petri-nets in order to model “nets within nets” [Va04]. The petri-net model was used to depict the material and energy flows associated with the activities of a scientific conference. We adapted [HH02] classification of environmental relevant conference activities, where the emissions of 3 main group activities are calculated via 12 subgroup activities as presented in Fig. .

[5] Umberto NXT Universal Version 7.1, more information at http://www.umberto.eu/
[6] Leading Life Cycle Inventory Database, more information at http://www.ecoinvent.org/database/

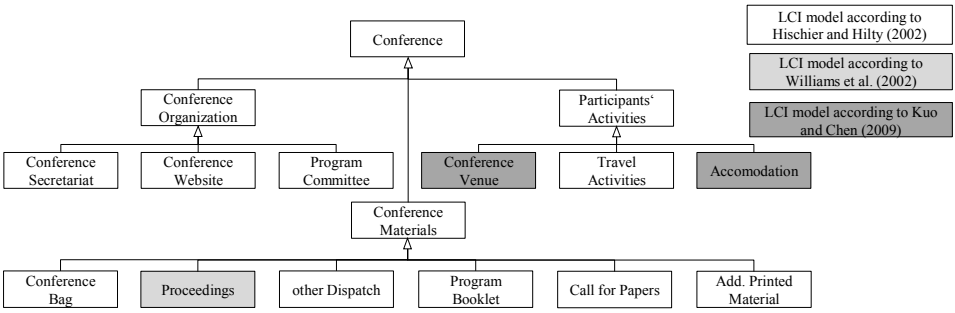


Fig. 1: Main model of the environmental relevant activities associated to a conference

For each of the 12 subgroup activities, we implemented material and energy flow models. Each subgroup activity model consists of one conference-specific activity as well as generic downstream activities. The generic downstream activities contain lists of material and energy flows (input and output) that are ascribed to a specific amount of product or service, e.g. 1 *km freight transport or 1 kg paper. The conference-specific activity contains the quantities of products and services that are used for one individual conference. The starting point for the implementation of the subgroup activities was the work of [HH02]. However, since [HH02] excluded the environmental impact of accommodation and conference venue, we extended the model by the results of [St08], [KC09] and [AAM13]. Moreover, we updated the material and energy flow model of the *proceedings* subgroup activity. The model by [HH02] includes a paper-based publication of conference proceedings, which was not the case for the last years IS conferences (Informatik, MKWI, WI, ECIS, ICIS). Therefore, the model was updated to publication by .pdf documents and USB devices. The material and energy flow model of the *proceedings* subgroup activity was based on the work of [WAH02]. For the parameterization of the model, we added data to the conference-specific activities as well as the generic downstream activities within the scope of 3 scenarios: A baseline scenario for a conventional conference in Central Europe, a remote location scenario for a Conference with long travelling distances and a dematerialization scenario by which conference materials are reduced to a minimum. In the baseline scenario, data from [HH02], [St08], [KC09], [AAM13] and [WAH02], was used to parameterize the conference-specific activities. Data within the literature was gathered for a 3-day conference in Central Europe with 308 participants. The data was linear extrapolated to an estimated number of 800 participants including 2400 overnight stays at a hotel in order to depict an average IS conference. Different transport systems were be estimated depending on the travelled distance d as shown in Tab. 3 [HH02].

Travelled Distance (one Way)	Amount Car	Amount Train	Amount Airplane
d < 300 km	10 %	90%	0 %
300 km < d > 1000 km	0 %	50 %	50 %
d > 1000 km	0 %	0 %	100 %

Tab. 3: Estimation for travel distances [HH02]

In contrast to existing works on the environmental impact of science [AAM13, CHB12, HH02, MC07, SL13, St08] we used data from the latest version of the ecoinvent database, ecoinvent 3.1, to parameterize the generic downstream activities. For the remote location scenario, we manipulated the travelling and accommodation data of the baseline scenario. The baseline scenario assumes a conference in a Central European country with a moderate climate and short travelling distances. As opposed to this, the remote location scenario depicts a conference with longer air travels and a warmer, subtropical climate at the conference venue. Tab. 4 shows the changes within the parameters compared to the baseline scenario. According to [KC09], the accommodation related wastewater discharge and water consumption level are doubled in order to model a warmer climate. Moreover, the datasets for the generic downstream activities are changed from European average datasets to Central and South American datasets. In the dematerialization scenario, the use of materials was reduced to a minimum. All information (call for papers, program booklet, etc.) is shared via email or homepage. Conference proceedings are provided solely online, no paper prints are used during the conference meetings and a conference bag is no longer needed. Additional energy consumption due the increased use of PC's, laptops and mobile devices for accessing the information have been neglected due to the results of [GK02]. An overview of all parameters for the conference-specific activities, the related ecoinvent downstream activities of the baseline scenario as well as examples for the petri-net models are available online[7] or available on request.

Parameter	Baseline Scenario	Remote Location Scenario
Amount of Car Travellers	5,5%	0,5 %
Amount of Train Travellers	58,6%	5,0 %
Amount of Short Distance Airplane Travellers	19,9%	20,0 %
Amount of Medium Distance Airplane Travellers	9,3%	24,5 %
Amount of Long Distance Airplane Travellers	5,8%	50,0 %
Water Consumption per Person and Night	292 L	584 L
Wastewater Discharge per Person and Night	200 L	400 L

Tab. 4: Estimation for the use of transportation [HH02]

4 Results and Discussion

The ReCiPe impact assessment method[8] was applied using the impact assessment engine of the Umberto software. Fig. shows the results of the impact assessment including 3 different metrics: greenhouse gas emissions measured in kilograms CO_2-equivalent (kg CO_2 eq), emission of environmental toxines (terrestrial ecotoxicity) measured in grams para-dichlorobenzene equivalent (g para-dichlorobenzene eq) and water depletion measured in cubic meters (m^3). More metrics (natural land transformation, urban land occupation, particular matter formation, etc.) were applied within the conduct of this

[7] http://bit.ly/1HJLTG1
[8] ReCiPe methodology for Life Cycle Impact Assessment, more information at http://www.lcia-recipe.net/

work but did not provide any additional information concerning the underlying research questions. The results of the baseline scenario are used to estimate the environmental impact and identify its main determinants. GHG emissions of around 211 t CO_2 eq. are estimated. This amount can be compared to the amount of CO_2 that is stored by woodland in the size of around 211 football fields in one year [BM03]. The majority of these emissions is determined by the conference related travelling activities which echoes the findings in recent work [AAM13, CHB12, HH02, MC07, SL13, St08]. However, emissions related to hotel stays and conference venue also account for more than third of the total GHG emissions which is significantly more than the literature suggests.

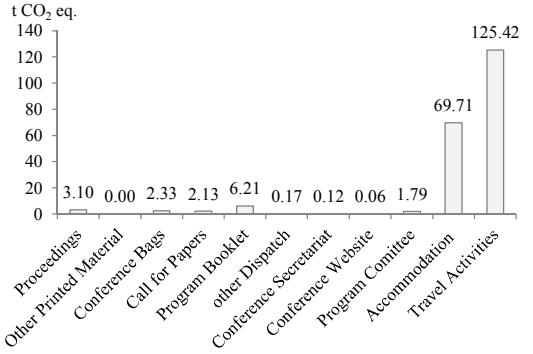


Fig. 2: GHG emissions of a conference (baseline scenario)

The use of additional metrics draws an even more differentiated picture as depicted in Fig. 3. Toxic substances equal to 30 kg para-dichlorobenzene are emitted into the earth's ecosystems. This amount can be compared to a lethal dose for 21 adults [Au14]. 47 % of the toxin emissions can be traced back to the use of pesticides and chemicals for cultivation and processing of cotton to produce conference bags. The amount of water that is used for a conference (1 835 m^3) would be enough to fill three quarters of an Olympic swimming pool. Almost the entire water consumption is determined by hotels stays and the use of conference venues. These results contradict those of [HH02], who argue that the choice of metric does not significantly affect the decision-relevant aspects of an environmental impact assessments results. Rather, the results suggest that the use of different metrics can lead to different focus areas to reduce environmental impact. The remote location scenario and the dematerialization scenario illustrate how the choice of the conference location and the organization of the conference can be used as effective instruments to reduce the environmental impact of a conference. The scenarios also draw a more complex picture of conference related environmental problems compared to [AAM13, CHB12, HH02, MC07, SL13, St08]. Fig. shows the results of all scenarios.

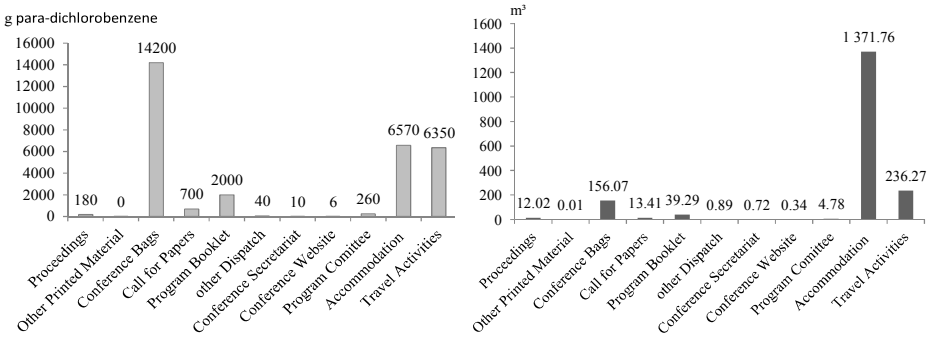


Fig. 3: Emission of toxic substances and depletion of water (baseline scenario)

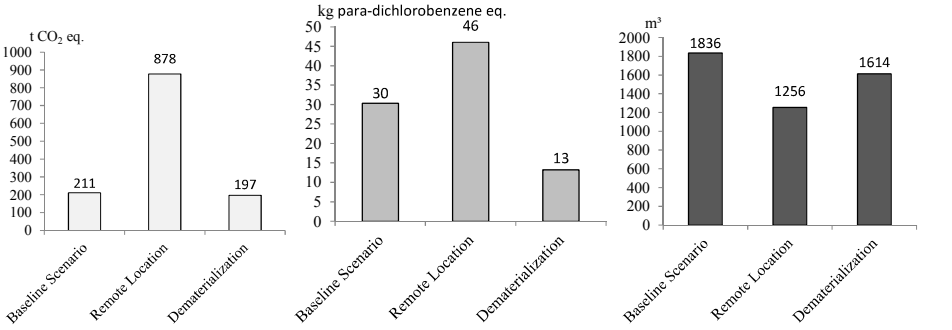


Fig. 4: Comparison of the three scenarios

The remote location scenario shows the consequences of a conference location requiring significantly more participants to take intercontinental flights. As a consequence, the GHG emissions increase by about 415 % and the emission of toxic substances by about 150 %. Besides the longer traveling distances, the increase can be explained by different types of electricity generation. Whereas the Central European energy mix contains huge amounts of nuclear and renewable energy sources, the global average energy mix is mainly determined by fossil energy sources [Ma10]. An unexpected result of the remote location scenario is the decrease of water depletion from 1 835 m³ to 1 255 m³ (31.6 %) per conference. A more detailed view on the calculation results reveals that the savings are mainly determined by more water-saving waste disposal in the global average data compared to the Central European data. This can be explained by a higher share of waste incineration in Central Europe, which requires significantly more water than a landfill [CBU09]. The dematerialization scenario reveals just a slight decrease in GHG emissions (about 6.6 %) and water depletion (12.1 %) due to diminished paper consumptions. GHG and water savings are quiet low because the major part of paper consumptions for a conference includes private printouts, those done by reviewers and authors at home. However, the dematerialization scenario shows a sharp decrease in terrestrial ecotoxicity (56.5 %) because cotton containing conference bags are not

necessary.

The results reveal implications for the scientific discourse on sustainable academic practicing. Three main determinants for the environmental footprint of major IS conferences have been identified: travel activities, accommodation and conference material. Effective instruments for organizers are accessing route minimization (e. g. by avoiding remote locations), reducing the number of hotel stays (e.g. by tightening the conference program, setting more conference events in parallel and arrive/depart straight after/before the conference) and avoid an extensive amount of conference material (e. g. share information online to avoid the use conference bags, print outs and booklets).

Due to the methodological nature of this work there are several limitations. The use of life cycle models requires idealizations and approximations that may be applied differently by different people. Therefore, the use of data and implementation of models are carefully documented. Another limitation results from the limited availability of empirical data about scientific conferences. We used linear extrapolation in order to apply data from the literature within our models. This however makes the model unsuitable to identify potential economics of scale or to determine an exact environmental footprint. Additionally energy consumption due the increased use of IT infrastructures for accessing the information have been neglected due to the results of [GK02]. As an additional scenario, a full video-conference scenario could complement the scenarios presented in this work. More research is therefore needed to implement more realistic and detailed models and collect empirical data.

5 Conclusion and Future Work

This work presents a life cycle assessment approach to understand more about the environmental impact of academic practicing. We suggest that measures with a focus on travel activities, accommodation and conference material are the most effective instruments to reduce the environmental footprint of major IS conferences. As a next step we will address the limitations of this work. Empirical data will be collected during a major IS conference in 2015 by using surveys and interviews. This data will be used to refine the presented model and improve the parameters in order to derive more reliable quantitative data.

Acknowledgement: This work is part of the project “Sustainable Consumption of Information and Communication Technologies in the Digital Society − Dialogue and Transformation through open innovation”. The project is funded by the Ministry for Science and Culture of Lower Saxony and the Volkswagen Foundation (VolkswagenStiftung) through the “Niedersächsisches Vorab” grant programme (grant number VWZN3037).

References

[AAM13] Achten, W.M.J.; Almeida, J.; Muys, B.: Carbon footprint of science: More than flying. In Ecological Indicators 34, S. 352–355, Nov. 2013.

[Au14] Australia New Zealand Food Standards Code: Extraneous Residue Limit (Para-Dichlorobenzene). 2014.

[BM03] Broadmeadow, M.; Matthews, R.: Forests , Carbon and Climate Change : the UK Contribution. Edinburgh, 2003.

[Bö06] Börjeson, L.; Höjer, M.; Dreborg, K.-H.; Ekvall, T.; Finnveden, G.: Scenario types and techniques: Towards a user's guide. In Futures 38/7, S. 723–739, Sep. 2006.

[Br09] vom Brocke, J.; Simons, A.; Niehaven, B.; Riemer, K.; Plattfaut, R.; Cleven, A.: Reconstructing the giant: on the importance of rigour in documenting the literature search process. European Conference on Information Systems. 2009.

[Br11] Brusco, J.M.: Trending toward paperless. In AORN journal 94/1, S. 13–8, Jul. 2011.

[Br13] vom Brocke, J.; Watson, R.T.; Dwyer, C.; Elliot, S.; Melville, N.: Green information systems: directives for the is discipline. In Communications of the AIS 33/30, S. 509–520, 2013.

[CBU09] Cherubini, F.; Bargigli, S.; Ulgiati, S.: Life cycle assessment (LCA) of waste management strategies: Landfilling, sorting plant and incineration. In Energy 34/12, S. 2116–2123, 2009.

[CHB12] Coroama, V.C.; Hilty, L.M.; Birtel, M.: Effects of Internet-based multiple-site conferences on greenhouse gas emissions. In Telematics and Informatics 29/4, S. 362–374, Nov. 2012.

[CLM11] Cellura, M.; Longo, S.; Mistretta, M.: Sensitivity analysis to quantify uncertainty in Life Cycle Assessment: The case study of an Italian tile. In Renewable and Sustainable Energy Reviews 15/9, S. 4697–4705, Dec. 2011.

[Co88] Cooper, H.M.: Organizing knowledge syntheses: A taxonomy of literature reviews. Knowledge in Society. 1988.

[Fi09] Finnveden, G.; Hauschild, M.Z.; Ekvall, T.; Guinée, J.; Heijungs, R.; Hellweg, S.; Koehler, A.; Pennington, D.; Suh, S.: Recent developments in Life Cycle Assessment. In Journal of Environmental Management 91/1, S. 1–21, Oct. 2009.

[GK02] Gard, D.; Keoleian, G.: Digital versus print: Energy performance in the selection and use of scholarly journals. In Journal of Industrial Ecology 6/2, S. 115–132, 2002.

[Gu02] Guinée, J.B.; Gorrée, M.; Heijungs, R.; Huppes, G.; Kleijn, R.; Koning, A. de; Oers, L. van; Wegener Sleeswijk, A.; Suh, S.; Udo de Haes, H.A.; Bruijn, H. de; Duin, R. van; Huijbregts, M.A.J.: Handbook on Life Cycle Assessment - Operational Guide to the ISO Standards. Kluwer Academic Publishers, Dordrecht, 2002.

[Ha96] Hartmann, S.: The World as a Process : Simulations in the Natural and Social Sciences. In (Hegselmann, R., Mueller, U., and Troitzsch, K.G., Hrsg.): Modeling and simulation in the social sciences from the philosophy of science point of view. Kluwer Academic Publishers, Dordrecht, S. 77–100, 1996.

[HC12] Hovorka, D.; Corbett, J.: IS sustainability research: A transdisciplinary framework for a "grand challenge." ICIS. S. 1–18, 2012.

[He96] Heijungs, R.: Identification of key issues for further investigation in improving the reliability of life-cycle assessments. In Journal of Cleaner Production 4/3–4, S. 159–166, Jan. 1996.

[HH02] Hischier, R.; Hilty, L.: Environmental impacts of an international conference. In Environmental Impact Assessment Review 22, S. 543–557, 2002.

[Is09a] ISO: Environmental management - Life cycle assessment - Principles and framework (ISO 14040), vol. 2006. International Organization for Standardization, 2009.

[Is09b] ISO: Environmental management - Life cycle assessment - Requirements and Guideline (ISO 14044). International Organization for Standardization, 2009.

[KC09] Kuo, N.-W.; Chen, P.-H.: Quantifying energy use, carbon dioxide emission, and other

environmental loads from island tourism based on a life cycle assessment approach. In Journal of Cleaner Production 17/15, S. 1324–1330, Oct. 2009.

[KG09] Klöpffer, W.; Grahl, B.: Ökobilanz (LCA): Ein Leitfaden für Ausbildung und Beruf, 1st ed. Wiley-VCH, Weinheim, 2009.

[Ma10] Marrero, G.A.: Greenhouse gases emissions, growth and the energy mix in Europe. In Energy Economics 32/6, S. 1356–1363, Nov. 2010.

[MBF11] Moberg, Å.; Borggren, C.; Finnveden, G.: Books from an environmental perspective - Part 2: e-books as an alternative to paper books. In The International Journal of Life Cycle Assessment 16/3, S. 238–246, Feb. 2011.

[MC07] Mathews, C.; Coetzee, D.: Reducing the carbon footprint of medical conference. In BMJ (Clinical research ed.) 334/7589, S. 323–326, Feb. 2007.

[Mi00] Ministry of Housing, S.P. and the E.: Eco-indicator 99 Manual for Designers - A damage oriented method for Life Cycle Impact Assessment, no. October. 2000.

[Mo10] Moberg, Å.; Johansson, M.; Finnveden, G.; Jonsson, A.: Printed and tablet e-paper newspaper from an environmental perspective — A screening life cycle assessment. In Environmental Impact Assessment Review 30/3, S. 177–191, Apr. 2010.

[MRS13] Mirabella, N.; Rigamonti, L.; Scalbi, S.: Life cycle assessment of Information and Communication Technology application: a case study of dematerialization in the Italian Public Administration. In Journal of Cleaner Production 44, S. 115–122, Apr. 2013.

[Mu89] Murata, T.: Petri nets: Properties, analysis and applications. In Proceedings of the IEEE 77/4, S. 541–580, 1989.

[NHE11] Niemann, C.; Hudert, S.; Eymann, T.: ON COMPUTER SIMULATION AS A COMPONENT IN INFORMATION SYSTEMS RESEARCH. 2011.

[OH02] O'Donnell, D.; Henriksen, L.B.: Philosophical foundations for a critical evaluation of the social impact of ICT. In Journal of Information Technology 17/2, S. 89–99, Jun. 2002.

[SL13] Spinellis, D.; Louridas, P.: The Carbon Footprint of Conference Papers. In PloS one 8/6, 2013.

[St08] Stohl, A.: The travel-related carbon dioxide emissions of atmospheric researchers. In Athmospheric Chemistry and Physics 8, S. 6499–6504, 2008.

[TP10] Tenhunen, M.; Penttinen, E.: Assessing the Carbon Footprint of Paper vs . Electronic Invoicing. 21st Australasian Conference on Information Systems (ACIS). Bristbane, 2010.

[Va04] Valk, R.: Object Petri-Nets. In (Desel, J., Reisig, W., and Rozenberg, G., Hrsg.): Lectures on Concurrency and Petri Nets. vol. 3098Springer Berlin Heidelberg, Berlin, Heidelberg, S. 819–848, 2004.

[WAH02] Williams, E.D.; Ayres, R.U.; Heller, M.: The 1.7 Kilogram Microchip: Energy and Material Use in the Production of Semiconductor Devices. In Environmental Science & Technology 36/24, S. 5504–5510, Dec. 2002.

[Wc87] WCED: World Comission on Environment and Development. Our Common Future. Oxford University Press, Oxford, 1987.

[We72] Weizenbaum, J.: On the impact of the computer on society. Science 176/4035, S. 609–614, 1972.

[WW02] Webster, J.; Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. In MIS Quarterly 26/2, S. 13–23, 2002.

[ZZ13] Zhang, L.; Zhang, Y.: A Comparative Study of Environmental Impacts of Two Delivery Systems in the Business-to-Customer Book Retail Sector. In Journal of Industrial Ecology 17/3, S. 407–417, Jun. 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Kommunikationsmöglichkeiten für Aspekte der Nachhaltigkeitsinformatik

Eva Kern[1]

Abstract: Umweltaspekte und eine nachhaltige Entwicklung gewinnen in der Gesellschaft zunehmend an Bedeutung. Zur Verbreitung der Thematiken können Informations- und Kommunikationstechnologien einen unterstützenden Beitrag leisten. Gleichzeitig hat die wachsende Digitalisierung aber auch negative Einflüsse auf Umwelt und Nachhaltigkeit, sowohl auf Hardware- als auch auf Softwareseite. Insbesondere letzteres ist auf Anwenderseite bisher wenig bis gar nicht bekannt, obgleich der Themenbereich der Umwelt- und Nachhaltigkeitsinformatik mit seinem Teilgebiet nachhaltiger Software in der Wissenschaft angekommen ist. Daher stellt sich die Frage, wie die Thematik verstärkt über wissenschaftliche Ansätze hinaus kommuniziert werden kann. In diesem Beitrag wird die Bedeutung der Kommunikation für die Umwelt- und Nachhaltigkeitsinformatik diskutiert. Anschließend werden erste Möglichkeiten zur Verbreitung von Aspekten grüner und nachhaltiger Software angerissen.

Keywords: Kommunikationsstrategie, Nachhaltigkeitskommunikation, Umweltinformatik, Nachhaltigkeitsinformatik, Green IT, Umweltinformationen

1 Einleitung

Die Themen „Umweltschutz“ und „Nachhaltigkeit“ gewinnen in der Gesellschaft zunehmend an Bedeutung. Mehr und mehr werden auch Fragen des Zusammenhangs zwischen Informations- und Kommunikationstechnik (IKT) und Umweltaspekten behandelt. Dabei ist klar, dass IKT sich sowohl positiv als auch negativ auf eine nachhaltige Entwicklung auswirken kann. Einerseits kann sie die Ressourcenschonung unterstützen („Green by IT“), andererseits verbraucht sie selbst über den kompletten Lebenszyklus Ressourcen. Es gilt, die durch IKT verursachten negativen Auswirkungen zu reduzieren („Green IT“). Hier spielen Kommunikation der Thematik und Transparenz im adressierten Feld eine große Rolle. Insbesondere im Softwarebereich fehlt es bisher weitestgehend an etablierten Lösungen, das Thema auf Anwenderebene anzugehen.

Auf Forschungsebene ist das Thema nachhaltiger Software seit ein paar Jahren präsent. Auf Konferenzen wie der „EnviroInfo“ oder der „ICT4S - Information and Communication Technologies for Sustainable Software“ wird das Thema in seinen zahlreichen Facetten diskutiert und weiterentwickelt. Darüber hinaus geschieht der Austausch auf Workshops: „Energy Aware Software-Engineering and Development (EASED)“, „Green and Sustainable Software (GREENS)“, „Software Engineering Aspects of Green Computing (SEGC)“, um nur einige Beispiel zu nennen, die in den letzten Jahren entstanden

[1] Leuphana Universität Lüneburg, Scharnhorststr 1, 21335 Lüneburg, mail@nachhaltige-medien.de

sind. Entsprechend finden sich unterschiedlichste Aspekte nachhaltiger Software und ihrer Entwicklung in Konferenzbeiträgen, Journals und zugehörigen Special Issues wie „Sustainable Computing: Informatics and Systems“. Wissenschaftler der Umweltinformatik und verwandter Disziplinen kommunizieren untereinander, nicht aber nach außen. Dieses Verhalten stellt Adomßent [Ad05] als typisch für die Wissenschaftskommunikation heraus. Der Abgeschlossenheit der Wissenschaft wirken öffentliche Vorträge wie bspw. im Rahmen einer „Nacht der Wissenschaft“ oder „Ringvorlesungen“ entgegen. Auf diese Weise werden wissenschaftliche Thematiken bewusst einer breiten Masse präsentiert und damit zugänglich gemacht.

Besondere Bedeutung erfährt diese Öffnung der Wissenschaft bei gesellschaftlich relevanten Themen wie dem Umweltschutz. Hier wird davon ausgegangen, dass die Effekte in der Umsetzung von wissenschaftlichen Erkenntnissen durch Mitwirken einer breiten Masse verstärkt werden können. Daher gilt es, auch das Thema nachhaltiger Informatik bekannter zu machen. Durch verschiedene Initiativen wie z. B. Produktkennzeichnungen und Veranstaltungen (bspw. war das Thema der CeBIT 2008 „Green IT“) kann zumindest im Hardwarebereich davon ausgegangen werden, dass ein Bewusstsein für die Thematik bei umweltinteressierten Personen vorhanden ist. Die Rolle von Software im Rahmen von Green IT scheint noch weitestgehend unbekannt.

Im Sinne des Workshop-Titels „Umweltinformatik zwischen Nachhaltigkeit und Wandel“ wird im folgenden Beitrag die Frage aufgeworfen, wie ein Wandel im Rahmen der Umwelt- und Nachhaltigkeitsinformatik als ein Wandel von einem „Wissenschaftswissen“ hin zu „Allgemeinwissen“ aussehen kann. Es wird die These vertreten, dass ein langfristiger Wandel hin zu einer nachhaltigen Entwicklung von einer großen Masse getragen werden sollte.

Die Nachhaltigkeitsinformatik wird dabei basierend auf Naumann [Na07] als „Teilgebiet der angewandten Informatik, welches sich mit der Erforschung und Entwicklung von Methoden, Modellen und Informationssystemen auseinander setzt, die zum Ziel nachhaltiger Entwicklung beitragen“ verstanden. Im Beitrag wird der darin enthaltene Aspekt der Umweltinformatik besonders betont („Umwelt- und Nachhaltigkeits-informatik“), da der Fokus der Betrachtung auf den ökologischen Aspekt gelegt wird.

2 Zielsetzung & Vorgehensweise

Das vorliegende Paper diskutiert Möglichkeiten der Nachhaltigkeitskommunikation, die zur Verbreitung des Themas nachhaltiger Informatik dienen können. Damit werden die folgenden Fragestellungen adressiert: Wie werden Aspekte der Umwelt- und Nachhaltigkeitsinformatik (bisher) kommuniziert? Welche Möglichkeiten lassen sich daraus für die Verbreitung der Bedeutung des Softwarebereichs auf dem Weg hin zu einer nachhaltigen Entwicklung ableiten? Ziel des Beitrags ist es, die Relevanz der Kommunikation für die Umwelt- und Nachhaltigkeitsinformatik aufzuzeigen Eine Evaluation der identifizierten Kommunikationsansätze wird nicht vorgenommen. Es handelt sich um ein Diskussionspapier, d. h. es besteht nicht der Anspruch die einzelnen Aspekte vollständig

und detailliert abzubilden. Vielmehr wird ein erster Überblick gegeben, der als Anstoß zur weiteren Bearbeitung des Themas dienen soll.

Die hier aufgeworfene Fragestellung ist dem Erkenntnisinteresse an Kommunikationsmöglichkeiten für den Bereich grüner Software zuzuordnen. Es wird davon ausgegangen, dass zur Erreichung des langfristigen Ziels einer Entwicklung von Bewertungsgrundlagen für grüne Software zunächst ein Bewusstsein für die Thematik vorhanden sein muss, bevor diejenigen Aspekte identifiziert werden können, die ein besonders hohes gesellschaftliches Interesse wecken.

Das Thema ist an der Schnittstelle der Themenfelder Kommunikation und Informatik einzuordnen. Daher wird in Abschnitt 3 allgemein auf Nachhaltigkeitskommunikation und ihren Zusammenhang zur Informatik eingegangen. Im darauf folgenden Abschnitt 4 werden Kommunikationsstrategien im Bereich der Nachhaltigkeitsinformatik vorgestellt. Basierend darauf werden Möglichkeiten zur Verbreitung der Thematik „nachhaltige Software“ für die Allgemeinheit vorgeschlagen. Abschließend werden in Abschnitt 5 die Ergebnisse zusammengefasst und nächste Schritte genannt.

3 Kommunikation als Voraussetzung für Relevanz

“Es mögen Fische sterben oder Menschen, das Baden in Seen oder Flüssen mag Krankheiten erzeugen, es mag kein Öl mehr aus den Pumpen kommen und die Durchschnittstemperaturen mögen sinken oder steigen: solange darüber nicht kommuniziert wird, hat dies keine gesellschaftlichen Auswirkungen.“ [Lu86]

Dieses Zitat macht die zentrale Rolle der Kommunikation deutlich. Insbesondere komplexe Themen, wie das der Nachhaltigkeit, müssen kommuniziert werden, damit ihre Bedeutung für die Gesellschaft die notwendige Relevanz in dieser erlangt. Die Kommunikation von Umwelt- und Nachhaltigkeitsthemen dient in diesem Fall der Bekanntmachung und Steigerung der Transparenz als Grundlage für umweltbewusstes Handeln und zur Unterstützung einer nachhaltigen Entwicklung.

3.1 Formen der Nachhaltigkeitskommunikation

Nachhaltigkeitskommunikation im Allgemeinen kann als „Verständigungsprozess, in dem es um eine zukunftsgesicherte gesellschaftliche Entwicklung geht, in deren Mittelpunkt das Leitbild der Nachhaltigkeit steht“ [Mi05] verstanden werden. Kommuniziert werden Werte wie inter- und intragenerationelle Gerechtigkeit. Es geht insbesondere um Problemwahrnehmung und darum, Handlungsmöglichkeiten aufzuzeigen. [Mi05] Durch die Kommunikation der Thematik erfährt diese die notwendige gesellschaftliche Relevanz. Dies ist Voraussetzung für Veränderungsprozesse hin zu einer nachhaltigeren Lebensweise im Sinne des Brundtland Reports. Michelsen [Mi05] führt verschiedene Beispiele für Instrumente der Nachhaltigkeitskommunikation auf, die zur indirekten Steuerung nachhaltigkeitsbezogenen Verhaltens führen können. Im Folgenden liegt der

Fokus auf Maßnahmen, die Endnutzer adressieren und sowohl grüne als auch nachhaltige IT einbeziehen.

Bildung für eine nachhaltige Entwicklung
Die Weltdekade „Bildung für eine nachhaltige Entwicklung“ (BNE) wurde für 2005 bis 2014 von den Vereinten Nationen ausgerufen. Ziel war es, das Leitbild der nachhaltigen Entwicklung im Rahmen von unterschiedlichsten Bildungsprojekten zu vermitteln. Ab 2015 wird die Bildungsoffensive in einem Weltaktionsprogramm BNE fortgeführt. Projekte der Dekade sind auf dem zugehörigen Webportal[2] gelistet.

Aktionen und Aufklärungskampagnen
Aktionen zur Förderung des Umweltschutzes und einer nachhaltigen Entwicklung sind bspw. der jährliche „Tag der Umwelt“ (5. Juni) mit wechselnden Mottos, die Klima-Aktion „Earth Hour“ und zahlreiche weitere Veranstaltungen, Kampagnen, Informationsstände und Maßnahmen. Sie dienen dazu Aufmerksamkeit zu schaffen.

Berücksichtigung von Nachhaltigkeitsaspekten in Verbrauchertests
Zur Unterstützung einer nachhaltigen Entwicklung können die Verbraucher auch in Produkttests wie Stiftung Warentest, TÜV, Öko-Test auf entsprechende Aspekte aufmerksam gemacht werden. Diese fließen dann direkt durch festgelegte Kriterien, z. B. Umweltauswirkungen und Energieverbrauch, in die Bewertung ein.

Labelling
Die Ergebnisse von Verbrauchertests werden oft in Form von bestimmten Labels auf den entsprechenden Produkten kenntlich gemacht. Auf diese Weise wird auf die Testergebnisse und zugehörige Produkteigenschaften hingewiesen. Darüber hinaus existieren zahlreiche Kennzeichnungen, die insbesondere beim Kauf eine Entscheidungshilfe darstellen sollen und dazu Informationen zu Umweltaspekten des Nutzungsgegenstands bereitstellen. Blauer Engel, Euro-Blume, Öko-Tex-Standard 100, Europäisches Umweltzeichen, Fairtrade, EU-Bio-Logo sind nur einige Beispiele für Labelling im Umweltbereich. Die Normenreihe ISO 14000 umfasst Richtlinien zur umweltbezogenen Produktinformation auf freiwilliger Basis für Unternehmen. Sie unterscheidet dabei drei Typen: wertend vergleichbare Kennzeichnungen, herstellerspezifische Umweltaussagen, Ökobilanz-basierte Informationen über Umweltauswirkungen ohne Wertung.

Umweltberatung
Neben Aufklärungskampagnen für die große Masse finden zunehmend spezielle Beratungsseminare zu Umwelt- und Nachhaltigkeitsthemen statt. Die Teilnehmenden werden zu speziellen Fragestellungen geschult, wie z. B. Umweltmanagement, Energiesparen, nachhaltiges Bauen. Abgesehen von Workshops und Seminaren existieren Beratungsbü-

[2] http://www.bne-portal.de/

ros als Anlaufstellen für Fragen zu Umweltschutz und nachhaltiger Entwicklung. In vielen Kommunen, Städten und Unternehmen agieren Energieberater, Nachhaltigkeits- bzw. Umweltbeauftragte, die sich mit Themen wie Abfall, Gewässerschutz oder Immissionsschutz beschäftigen und diese aktiv kommunizieren.

3.2 Nachhaltigkeitskommunikation und Informatik

Bei der Betrachtung der Nachhaltigkeitskommunikation und der Nachhaltigkeitsinformatik lassen sich verschiedene Parallelen feststellen: Wie auch viele weitere Nachhaltigkeitsdisziplinen entwickelten sich beide aus dem Umweltbereich heraus. Das heißt, zunächst wurden Umweltthemen kommuniziert bzw. im Informatikbereich einbezogen, bis sich spätestens nach der Rio-Konferenz 1992 das Bewusstsein verbreitete, dass bei Umweltfragen das Prinzip einer nachhaltigen Entwicklung eine entscheidende Rolle spielt [Mi05]. Als Konsequenz daraus werden inzwischen auf der Kommunikationsebene und im IT-Sektor mindestens auch ökonomische und soziale Aspekte einbezogen. Die Bezeichnungen „Nachhaltigkeitskommunikation“ und „Nachhaltige Informationsgesellschaft“ [Do04] bzw. „Nachhaltigkeitsinformatik“ [Na07] haben sich in den Wissenschaftsbereichen etabliert. In beiden Disziplinen wird nach wie vor der Schwerpunkt auf den Ausgangspunkt, d. h. Fragen der Ökologie, gelegt. Darüber hinaus können in beiden Bereiche „in- und by-Aspekte“ gefunden werden: Die Unterscheidung zwischen Green IT und Green by IT ist auf „Nachhaltige IT“ und „Nachhaltig by IT“ übertragbar. Gleichermaßen kann eine nachhaltige Entwicklung durch Kommunikation gefördert bzw. oft sogar erst angestoßen werden („Nachhaltig by Kommunikation“). Andererseits kann auch von einer nachhaltigen Kommunikation gesprochen werden, wenn bspw. die Kommunikationsform aus ökologischer, ökonomischer und soziale Sicht bewertet wird.

Am Beispiel der Nachhaltigkeitskommunikation im Internet wird deutlich, inwiefern IKT dazu beitragen kann Nachhaltigkeitsthemen zu transportieren: Websites wie das Utopia-Portal[3], Weeyoo[4], und themenspezifische Gruppen, Foren und Profile in sozialen Netzwerken regen dazu an, sich mit Nachhaltigkeitsaspekten im Alltag auseinander zu setzen. Derartige Plattformen dienen insbesondere der Vernetzung und dem Austausch von gemeinsamen Interessen. Neben der Diskussion über Aktionen und Produkte für eine bessere Welt, verfolgen sie das Ziel einer Bewusstseinsänderung [Ho12].

Dieser Bewusstseinsänderung bedarf es auch im Bereich der Informatik selbst. Zahlen und Prognosen zum Energieverbrauch durch IKT (z. B. [Fi12]) machen deutlich, dass Handlungsbedarf in diesem Bereich besteht. Daher wird im Folgenden darauf eingegangen, welche Maßnahmen die Generierung eines Bewusstseins für die Relevanz der IKT für eine nachhaltige Entwicklung unterstützen können.

[3] http://www.utopia.de/
[4] http://www.weeyoo.de/

4 Kommunikation in und über Nachhaltigkeitsinformatik

Kommunikation über den Zusammenhang zwischen IT und nachhaltiger Entwicklung kann helfen Transparenz in diesem Feld zu generieren und ein Bewusstsein zu schaffen. Im folgenden Abschnitt wird zunächst ein Rahmen gebildet, um eine gemeinsame Grundlage zu schaffen (Wer kommuniziert über was mit wem?). Anschließend wird ein Überblick gegeben, welche Kommunikationsformen bisher im Bereich Umwelt- und Nachhaltigkeitsinformatik genutzt werden. Aufbauend darauf, werden Möglichkeiten zur Verbreitung der Bedeutung der Softwareseite der Green IT vorgeschlagen. Die Aspekte „Green by IT“ bzw. „Green by Software“ werden außen vor gelassen.

4.1 Rahmensetzung

Nachhaltigkeitskommunikation findet im Allgemeinen auf unterschiedlichen Ebenen und in verschiedenartigen Kontexten statt [Mi05]. Deshalb ist zunächst zu klären, welche Kommunikationspartner, Situationen und Gegebenheiten hinsichtlich der Kommunikation im Bereich Umwelt- und Nachhaltigkeitsinformatik adressiert werden. Dazu wird auf die Fragen der Kommunikationsstrategie nach Bruhn [Br05] zurückgegriffen.

- **Wer? (Kommunikationsobjekt)**: An der Kommunikation können Nutzer, Entwickler, Produzenten, Nachhaltigkeitsbeauftragte, aber auch Institutionen und bspw. Unternehmen beteiligt sein. Penzenstadler et al. [Pe13] identifizieren Stakeholder für den Nachhaltigkeitsbereich. Diese unterscheiden sich je nach Kontext. Ihnen gemeinsam ist, dass sie bekannt sein müssen, um eine nachhaltige Entwicklung unterstützen zu können. Im vorliegenden Beitrag werden in erster Linie Unternehmen der IKT-Branche und unabhängige Institutionen (z. B. Beraterbüros, Kompetenzzentren und ähnliche Einrichtungen) als Kommunikationsobjekte angesehen.
- **Wem? (Kommunikationszielgruppen)**: Ebenso wie die Kommunikationsobjekte sind auch die Zielgruppen unterschiedlich. Hier werden vorwiegend Möglichkeiten vorgestellt, die insbesondere die Gruppe der Endnutzer (Anwender) adressieren. Diese Gruppe spielt vor allem im Kontext mit einer energiesparenden IKT eine große Rolle, da in der Nutzungsphase ein hoher Energieverbrauch entsteht, der durch den Nutzer verantwortungsbewusst gesteuert werden kann, vorausgesetzt er ist darüber informiert. Im Green House Gas Protocol [Ge13] wird in diesem Zusammenhang insbesondere die Bedeutung der Software während der Nutzungsphase hervorgehoben. Die Studie geht davon aus, dass bis zu 90 % der durch Hardware genutzten Energie durch die auf der Hardware ausgeführte Software zurückzuführen ist.
- **Was? (Kommunikationsbotschaft)**: Die zu übermittelnde Nachricht umfasst Informationen aus dem Bereich der Umwelt- und Nachhaltigkeitsinformatik. Je nach Kommunikationsobjekt und Kommunikationszielgruppe werden unterschiedliche Arten von Botschaften vermittelt. Diese können vergleichend (z. B. „In

Energiemessungen hat Produkt A weniger Energie verbraucht als Produkt B“), informierend (z. B. „Software verbraucht Energie“), belehrend (z. B. „Ziehen Sie Ladegeräte nach dem Aufladen aus der Steckdose“) oder auch auffordernd (z. B. „Reduzieren Sie den Energieverbrauch ihrer IT-Geräte“) sein.

- **Wie? (Kommunikationsmaßnahmen)**: In Abschnitt 3.1 wurde auf die Frage der Kommunikationsform mögliche Antworten gegeben. Diese werden in den folgenden Abschnitten um Beispiele aus dem Green IT Bereich und Ideen zur Kommunikation von Aspekten grüner und nachhaltiger Software ergänzt. Die Auflistungen erheben keinen Anspruch auf Vollständigkeit.

Die Fragen nach dem **Wo?** (Kommunikationsareal) und **Wann?** (Kommunikationstiming) werden hier vernachlässigt, da diese sich je nach Kommunikationsmaßnahme stark unterscheiden können. Sie sind festzulegen, nachdem die Maßnahmen, mit der Aspekte der Umwelt- und Nachhaltigkeitsinformatik kommuniziert werden sollen, festgelegt sind. Grundsätzlich gilt, dass die Kommunikation möglichst überregional und kontinuierlich geschehen sollte, um einen großen Effekt zu erreichen.

4.2 Stand der Kommunikation im Nachhaltigkeitsinformatik-Bereich

Die vorgestellten Beispiele für existierende Kommunikationsformen im Green IT Bereich sind Ergebnisse einer Webanalyse. Die analysierten Formen orientieren sich an den in Abschnitt 3.1 vorgestellten Beispielen. Die Beispiele sind als einzelne Spots zu sehen, die eine Idee vorhandener Ansätze vermitteln.

Bildung für eine nachhaltige Entwicklung

Im BNE-Bereich spielen auch Fragen der Green IT eine Rolle. Das zugehörige Portal listet verschiedene Projekte in diesem Themenfeld auf. Darüber hinaus werden auf der Kampagnen-Website Themen wie „Nachhaltigkeit in digitalen Spielen“, „Green IT - Mehr als nur Stromsparen“ und „(…) Green IT in Schule und Unterricht“ adressiert.

Aktionen und Aufklärungskampagnen

Zur Information über Themen der Green IT werden zunehmend Informationsbroschüren durch Institutionen und Initiativen wie die Deutsche Energie-Agentur (dena), dem Umweltbundesamt (UBA), Greenpeace oder „Make IT fair“ herausgegeben. Die Broschüre „Computer, Internet und Co“[5] ist nur eines von vielen Beispielen. Zur Unterstützung der frühzeitigen Bildung im Bereich Green IT stellt u. a. das UBA entsprechende Lehrmaterialien[6] bereit.

Berücksichtigung von Nachhaltigkeitsaspekten in Verbrauchertests

Während Nachhaltigkeitsaspekte nur langsam Beachtung in bekannten Verbrauchertests

[5] http://www.umweltbundesamt.de/sites/default/files/medien/publikation/long/3725.pdf
[6] http://www.umweltbundesamt.de/en/publikationen/green-it

finden, werden bspw. von Greenpeace Produkt- und Herstellerbewertungen („Guide to Greener Electronics“ und „Green Electronics Survey“) von IT-Firmen hinsichtlich Fragen eines nachhaltigen Engagements herausgegeben. Online-Portale wie die „WeGreen Nachhaltigkeitsampel“[7] beziehen auch die IT-Branche ein.

Labelling

Im Hardwarebereich können verschiedene Produktkennzeichnungen zur Kaufberatung herangezogen werden. Diese umfassen alle der in Abschnitt 3.1 beschriebenen Typen. Zu den bekanntesten Labels im Green IT Bereich zählen: Blauer Engel, Euro-Blume, ENERGY STAR, TCO-Gütesiegel, EU-Label, TÜV SÜD Gütesiegel.

Umweltberatung

Zur Vermittlung von Green IT Themen wurden und werden Workshops und Seminare angeboten, die bspw. Hinweise zu Möglichkeiten des Energiesparens im IT-Bereich geben oder über entsprechende Fördermöglichkeiten informieren. Veranstalter sind z. B. die IHK, das UBA, die dena und ähnliche Institutionen. Von 2009 bis 2012 förderte das Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit zudem das sogenannte Green IT Beratungsbüros beim BITKOM.

4.3 Mögliche Instrumente im Bereich „Grüne und nachhaltige Software“

Die vorgestellten Kommunikationsmaßnahmen beziehen insbesondere Green IT Aspekte der Hardware ein. Für den Software-Bereich sind derartige Informationsangebote, Handlungsempfehlungen, u. ä. bisher noch wenig bis gar nicht verfügbar. In Folge dessen, ist der Einbezug von Softwareaspekten in den Bereich der Umwelt- und Nachhaltigkeitsinformatik noch nicht verbreitet. Um dem entgegenzuwirken, werden im Folgenden Möglichkeiten angerissen, die als Grundlage für Konzepte und Instrumente zur Kommunikation des Themenfeldes grüner und nachhaltiger Software dienen können. Allgemeinen ist zu empfehlen die existierenden Maßnahmen im Green IT Bereich um Fragen der Softwareseite zu ergänzen. Der Vorteil dabei ist, dass diese bereits bekannt und verbreitet sind.

Bildung für eine nachhaltige Entwicklung

Im Rahmen der BNE-Dekade entstanden insbesondere auch zahlreiche kreative Projekte, die sich von bekannten Kommunikationsmaßnahmen unterscheiden. Besonderer Wert wurde auf Partizipation, kulturelle Vielfalt und internationale Verständigung gelegt. Vorstellbar wären hier Projekte wie ein Filmwettbewerb ähnlich zu GreenITube[8] jedoch mit explizitem Einbezug von Softwareaspekten, (internationale) Sommer- und/oder Schüler-Unis zu Themen nachhaltiger IKT, vergleichbare Workshops bspw. inklusive Energieverbrauchsmessungen von Software. Dabei sollte Bildung im Vordergrund ste-

[7] http://www.wegreen.de/
[8] https://www.izt.de/fileadmin/templates/projekte/GreenIT/Ausschreibung-greenITube_lang.pdf

hen, verbunden mit alternativen Lehr- und Lernmethoden für unterschiedliche Altersklassen.

Aktionen und Aufklärungskampagnen
Während Projekte im Bereich von BNE kreativer und informeller gestaltet sein können, sollten thematisch ähnliche Aufklärungskampagnen eher seriös gestaltet sein. Neben der Erweiterung existierender Materialien um Softwareaspekte, können eigene Informationsbroschüren dazu herausgegeben werden. Aufgrund des bisherigen Forschungsstands lassen sich noch keine direkten Handlungsempfehlungen zu Softwareprodukten geben. Es ist z. B. nicht möglich, sich aufgrund eines geringeren Energieverbrauchs für ein bestimmtes Produkt und gegen ein anderes auszusprechen. Diese Unklarheiten werden bspw. im Fall von Energieverbrauchsmessungen von Internet Browsern deutlich, in denen sich je nach Test unterschiedliche Produkte energiesparender zeigten [Vo11], [Ke11]. Es kann jedoch bereits kommuniziert werden, dass die Nutzung von unterschiedlicher Software einen Einfluss auf den ressultierenden Energieverbrauch hat.

Berücksichtigung von Nachhaltigkeitsaspekten in Verbrauchertests
Um Nachhaltigkeitsaspekte in Verbrauchertests einzubeziehen, sind entsprechende Kriterien zu entwickeln. Beispielsweise fehlt es an standardisierten Nutzungsszenarien zur Überprüfung des Energieverbrauchs von Software mit Hilfe von Energieverbrauchsmessungen. Darüber hinaus können Aspekte wie die Bewertung der Arbeitsbedingungen in Softwareunternehmen einbezogen werden. Auch das Softwareprodukt selbst kann hinsichtlich nachhaltiger Kriterien untersucht werden. Eine Idee ist z. B. die Frage nach dem Grad der Modularisierung der Software. Bisher existieren verschiedene Ansätze für mögliche Nachhaltigkeitskriterien für Software, die dann in Verbrauchertests berücksichtigt werden können; eine Einigung auf bestimmte Kriterien fehlt bislang.

Labelling
Auch in der Label-Frage gibt es erste Ideen und Fragestellungen, die aktuell angegangen werden. Es zeichnet sich jedoch noch keine klare Tendenz ab. Unklarheit herrscht insbesondere bezüglich einer einheitlichen Definition, zu Grunde liegender Kriterien und der Darstellungsform.

Umweltberatung
Bei vorhandenem Material können entsprechende Schulungen, Workshops und Seminare konzipiert werden, die sich bspw. an Entscheider im Unternehmen, Endanwender im privaten Umfeld aber auch direkt an Softwareentwickler wenden.

5 Zusammenfassung & Ausblick

Zusammenfassend kann festgehalten werden, dass die Frage der Kommunikation im

Bereich der Nachhaltigkeitsinformatik, einer noch jungen Wissenschaftsdisziplin, von hoher Relevanz ist. Insbesondere wenn das Ziel einer Transparenz für Anwender angestrebt wird, kann sie als Antreiber einer solchen Bewegung angesehen werden, denn die Umweltrelevanz insbesondere von Software hat den Verbraucher bisher nicht erreicht. Darüber hinausgehende Aspekte der Nachhaltigkeit sind zunächst mit entsprechenden wissenschaftlichen Fragestellungen zu analysieren, bevor sie zu gesellschaftlichen Themen ausgeweitet werden können. Entsprechend gilt es die vorgeschlagenen Möglichkeiten näher zu differenzieren, Konzepte zu entwickeln und anschließend in die Praxis umzusetzen, um damit eine große Reichweite zu erzielen.

Literaturverzeichnis

[Ad05] Adomßent, M.; Godemann, J.: Umwelt-, Risiko-, Wissenschafts-und Nachhaltigkeitskommunikation: Eine Verortung. In Handbuch Nachhaltigkeitskommu nikation. Grundlagen und Praxis. Oekom-Verlag, München, 2005; S. 43–52.

[Br05] Bruhn, M.: Kommunikationspolitik. Systematischer Einsatz der Kommunikation für Unternehmen. Vahlen, München, 2005.

[Do04] Dompke, M. et al.: Memorandum Nachhaltige Informationsgesellschaft. Fraunhofer-IRB-Verlag, 2004.

[Fi12] Fichter, K.; Hintemann, R.; Beucker, S.; Behrendt, S.: Gutachten zum Thema "Green IT - Nachhaltigkeit". für die Enquete-Kommission Internet und digitale Gesellschaft des Deutschen Bundestages.

[Ge13] GHG Protocol Product Life Cycle Accounting and Reporting Standard ICT Sector Guidance. Chapter 7, Software. Draft v2.9, 2013.

[Ho12] Hoffhaus, M.: Die „sieben Todsünden “der CSR-und Nachhaltigkeitskommunikation und wie ein nötiger Paradigmenwechsel im Verständnis von Kommunikation zu mehr Glaubwürdigkeit von Organisationen beitragen kann. In uwf-UmweltWirtschaftsForum, 2012; S. 155–163.

[Ke11] Kern, E. et al.: Green Software and Green IT: An End User Perspective. In (Golinska, P.; Fertsch, M.; Marx-Gómez, J. Hrsg.): Information Technologies in Environmental Engineering. Proceedings of the 5th International ICSC Symposium on Information Technologies in Environmental Engineering. Springer, Berlin, 2011; S. 199–211.

[Lu86] Luhmann, N.: Ökologisches Wissen und gesellschaftliche Kommunikation. Springer, 1986.

[Mi05] Michelsen, G.: Nachhaltigkeitskommunikation: Verständnis Entwicklung Perspektiven. In Handbuch Nachhaltigkeitskommunikation. Grundlagen und Praxis, 2005; S. 25–41.

[Na07] Naumann, S.: Nachhaltigkeitsinformatik. Ein neues Teilgebiet der angewandten Informatik? In: Geiger, W.; Schreiber, M. (Hrsg.), Rundbrief des Fachausschusses 4.6. Informatik im Umweltschutz, 2007; S. 9–12.

[Pe13] Penzenstadler, B.; Femmer, H.; Richardson, D.: Who is the advocate? Stakeholders for sustainability: Green and Sustainable Software (GREENS), 2013 2nd International Workshop on, 2013; S. 70–77.

[Vo11] VonKoch, W.; Robben, M.; Weber, J.: Browser Power Consumption - Leading the Industry with Internet Explorer 9.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

SALT – Situation-sensitive Sustainable Service and Product Alternatives: Vision, Conceptual Application & Challenges

Dominik Trumm[1], Andreas Filler[1], Eva Kern[2], Stefan Naumann[1]

Abstract: The field of environmental informatics can clearly contribute to the achievement of the Sustainability Goals set up by the German Government. To do so, it has to investigate not only environmental aspects but also integrate social and economic aspects. In this work-in-progress paper this will be done by proposing a system for visualizing **s**ustainable service and product **alt**ernatives (SALT). Starting from the observable gap between the intention to behave more sustainable and its transformation into sustainable behavior, it is elaborated how a SALT-system can help to bridge the gap in this domain. After introducing a conceptual architecture, the example application of a browser plugin for presenting lending alternatives to online shoppers is described.

Keywords: sustainable by ICT, service and product alternatives, intention-behavior gap, sustainable behavior

1 Introduction

In 2002, the German Government submitted the national strategy for Sustainability Goals of Germany, called "Perspectives for Germany – Our Strategy for Sustainable Development" [Fe02]. Thus, the German Government points out that its governance wants to follow the principles of a sustainable development on the one hand and reports about the progress regularly on the other hand. Until now, much effort to achieve the goals of 21 topics of sustainable development in Germany has been invested. Especially regarding the first goals addressing the saving of resources, climate protection, renewable energies, and information and communication technologies (ICT) can make a contribution to the strategy for sustainable development.

So far, the research of environmental informatics mainly focuses on the environmental aspects of sustainability. However, the movement should also take the coherencies and effects of ICT on the different dimensions of sustainability, i.e. environmental, economical, and social into account. Regarding software, one of the additional aspects

[1] University of Applied Science Trier, Environmental Campus Birkenfeld, Postfach1380, 55761 Birkenfeld, d.trumm@umwelt-campus.de, a.filler@umwelt-campus.de, s.naumann@umwelt-campus.de
[2] Leuphana Universität Lüneburg, Scharnhorststr. 1, 21335 Lüneburg, mail@nachhaltige-medien.de

next to the environmental impacts caused by software is the development of sustainable software products. Penzenstadler et al. [Pe14] define sustainable software as follows: "The term Sustainable Software can be interpreted in two ways: (1) the software code being sustainable, agnostic of purpose, or (2) the software purpose being to support sustainability goals, i.e. improving the sustainability of humankind on our planet. Ideally, both interpretations coincide in a software system that contributes to more sustainable living." Whereas (1) is commonly referred as Green (in) IT, (2) is known as Green by IT.

Hence, there are two directions within the research activities in the context of environmental and sustainable informatics that can support a strategy for sustainable development: approaches dedicated to Green (in) IT address the question of how software can be sustainable and how the sustainability of software can be measured (e.g. [Ab15], [ANC12]). Activities belonging to the Green by IT field aim to support a sustainable development by using ICT [Pe13]. If including more than just the environmental aspects of sustainability one can talk about "Sustainable (in) ICT" and "Sustainable by ICT".

The following article concentrates on "Sustainable by ICT". It deals with the problem of the divergence between the growing awareness for sustainability issues on the one hand and the missing sustainable behavior (the intention-behavior gap [AD07]) on the other hand (e.g. [Fu05]) and proposes a tool, which can support sustainable behavior.

In the next chapter, we outline different it-related approaches trying to foster sustainable behavior. Following we introduce a model that helps to explain the intention-behavior gap (Section 3) and argue why the presentation of **s**ustainable service and product **alt**ernatives (SALT) could help to bridge this gap (Section 4). Section 5 is dedicated to the introduction of a conceptual architecture for a SALT system and Section 6 shows an example application of such a system. Finally, we will conclude this work-in-progress article by a conclusion and consideration of future challenges in Section 7.

2 IT-approaches to foster sustainable behavior

The field of sustainable human-computer interaction is rich of approaches, which try to promote a change in behavior towards a more sustainable lifestyle [DSB10]. Ambient awareness systems visualize the impact of a user's behavior on particular aspects of sustainability (e.g. [Hu10]). Persuasive technologies [Fo02] go one step further and try to persuade users to act in a more sustainable way.

Although DiSalvo et al. [DSB10] count in its survey of sustainable human-computer interaction about 45% of all work to the subgenre of persuasive technologies, the question arises if persuasion is the right strategy to change behavior. In [HH15] several limitations of these approaches are listed, including the inherent technology paternalism and the insufficient account of individual differences and social context. As a way to

overcome these limitations, Hubel et al. [HH15] concentrate on gamification as a method to promote sustainable behavior and develop requirements for such types of systems. Examples of a system that combines gamification elements with the attempt to foster sustainable behavior are the WeAct Challenge [We15] and the Power Agent [GKB09]. Gustafsson et al. [GKB09] proof in his evaluation that gamification can improve the motivation to act more sustainable. However they could not show that gamification can lead to a long-term behavior change.

The approach presented in this paper concentrates not on motivational aspects or the distribution of general information about sustainability, but far more on the problem of converting intentions to behavior in the domain of sustainability.

3 Modeling the intention-behavior gap

An often used and well-known model for predicting behavior is the theory of planned behavior [Aj91]. In that theory perceived behavioral control, subjective norms and the attitude towards the behavior mediate between belief and intentions. It states that if the intention becomes stronger, a behavior is performed more likely. This assumption is challenged by research work that shows that intentions are a bad predictor for a behavior (e.g. [Sh02]). This problem is described in literature as the intention-behavior gap (e.g. [Fe11]).

Carrington et al. [CNW10] try to explain this gap in their intention-behavior mediation and moderation model of the ethically minded consumer (see Fig. 1). In their work they combine the three already existing concepts of actual behavioral control [AM86], implementation intentions [Go99], and situational context [Be75] to explain how a transition from an intention to the actual behavior can occur.

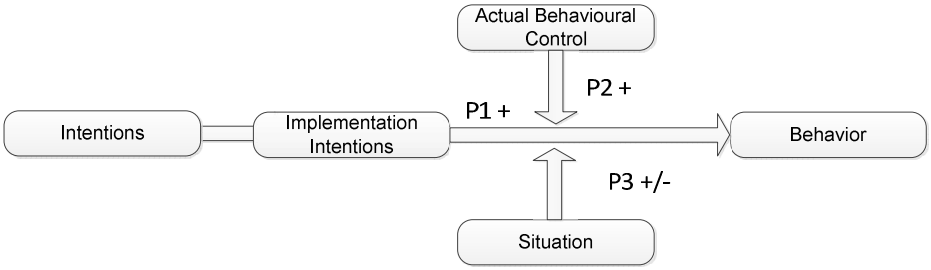


Fig. 1: Intention-behavior mediation and moderation model [CNW10]

Implementation intentions are plans how an intention can be transformed into a behavior. An example for such a plan for the intention “I want to buy more fair-trade clothes” can be formulated in an if/then statement like “If I need a new shirt and I am searching on the web then I will use the online shop of a fair trade clothing company”. Actual behavioral control describes the extent to which a person can control his own

actions in a particular situation. This construct includes constraints like time, resources or also knowledge, which could hinder a person from executing an implementation intention. The situational context (referred just as situation in this paper) can be described as the physical and situational environment at the time the intention should be transformed into a particular behavior. According to Belk [Be75], five factors describe a situation: Physical Surroundings, Social Surroundings, Temporal Perspective, Task Definition and Antecedent States.

While the intention-behavior gap is positively moderated by all of the three named constructs, it can be also moderated negatively in the case of the situational context [CNW10]. Examples for the negative influence of the situation, regarding the fulfillment of an intention, are special offers, which could influence people to deviate from their implementation intention.

Based on this model the next section is dedicated to the question why the visualization of service and product alternatives could be a useful tool to support sustainable behavior by reducing this gap.

4 Service and product alternatives and the intention-behavior gap

Presenting sustainable service and product alternatives to a user can – if meeting certain requirements – have a positive impact on all of the three aforementioned constructs.

First, implementation intentions can be supported by the presentation of alternatives in several ways. The alternative can be new to the user thus enabling him/her to form a new implementation intention. Existing plans can be improved or the complete planning process can be performed by such a system.

Through the visualization of alternatives, a situation can be shaped in favor to more sustainable behavior. The presentation, e.g. on a smartphone or on a wearable, can remind users of their intentions and work as a compensator for other visual temptations (e.g. special offers).

At last, applications following the approach presented in the next sections can improve the actual behavioral control over a particular situation. These applications are able to mitigate constraints like time or knowledge; as the knowledge lies in the system and the push structure of these applications supersede the need to actively search for alternatives.

For being successful in supporting users to change their behavior, the proposed alternatives have to be directly actionable, fitting to the situation and of a higher utility in regard to the “status-quo”-behavior. Directly actionable means that the “then” part of an implementation intention is communicated with the alternative (e.g. “go 100 meter further, there is a fair-trade shop for buying tea.”). The accordance of situation and alternative can be captured by the situation-service fitness. This concept is defined by Maass et al. [Ma12] as “a perceived degree of match between a service and a situation.

Following TTF [Task-technology fit theory], SSF [situation-service-fitness] theory hypotheses that the better the match between service and situation, the better the individual or group performance within a situation." The utility of the alternative relates to the effort to realize the alternative in comparison to given constraints or an induced saving. To exemplify the said: Taking a bus instead of driving a car is in general a better idea as it helps to reduce CO_2 emissions, but if the bus station is too far away for walking under given constraints (e.g. weather or a meeting at a defined time) the alternative is less probable to be chosen.

5 The SALT system

In this section a system is proposed, which enables the presentation of service and product alternatives, as elaborated in the former section. A conceptual architecture for the SALT system can be seen in Fig. 2.

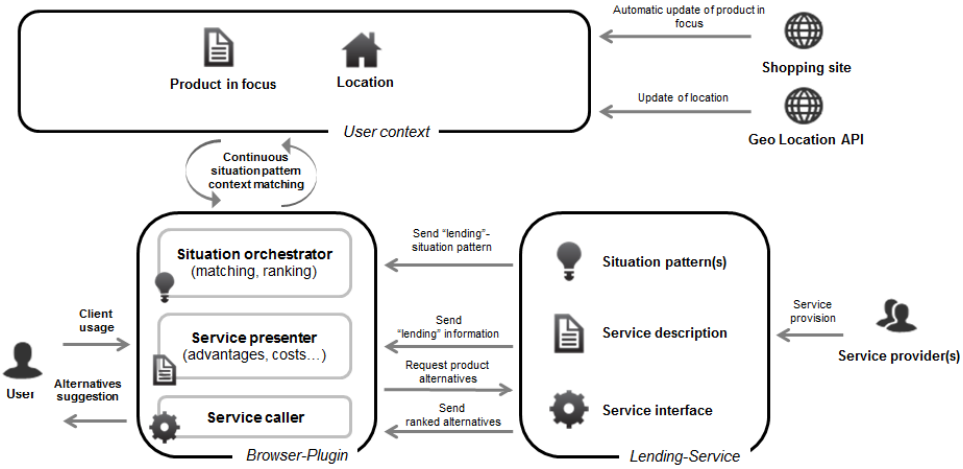


Fig. 2: Conceptual architecture of the SALT system

The starting point of such a system is the user, who is associated with a context. This context consists in conformity with Carrington et al. [CNW10] of the following information: A user profile, environmental information, a social context and revisions of this data over time.

The data describing the user context can be gained from the user itself in two ways: by manually informing the system about facts, preferences or personal goals or by automatically determining it, based on events triggered by user actions. Beside this way of gathering data many other available data sources can be used to describe the context and increase its accuracy. Among these are: basic information about time, location, weather, social information like contacts and appointments of/with friends, colleagues etc., e.g. retrieved from the smartphone of the user, IoT devices, social networks, and

more. While the context is always updated with the newest information state, it still holds former revisions over time to enable historical data analysis, e.g. to determine continuous changes, reoccurring patterns over time.

The management of such context-related data is performed by a not further elaborated technical infrastructure, which could be cloud-based or directly on the smartphone of the user, but that's not a focus of this publication.

Based on the user context, a user's personal client, which could be a smartphone app, a web application or also implemented in a wearable will support the process of providing alternatives using three major components: A situation orchestrator, a service presenter, and a service caller. The situation orchestrator is continuously matching the current and former context states with different situation patterns, which are aggregated from the services that want to provide alternatives to the user. These patterns describe situations (= subsets of a context [Ma12]) in which a particular service can possibly compute fitting product or service alternatives. If a pattern match is detected in this continuous matching process, the corresponding service is presented to the user using the service presenter and the retrieved services description. The service presenter is able to provide the information in an attractive way to the user, showing advantages of the service, service costs and more. At the same time, the service caller calls the service and provides the service context information to enable the calculation of alternatives. The service delivers a created list of personally calculated alternatives fitting to the context of the user back to the service caller to enable the presentation of the alternatives next to the general information about the service.

If the user decides to use one of the provided alternatives the service caller also manages the further communication between the client and the service, involving service delivery, payment and more.

The presented procedure requires the services to provide all information and interfaces to be able to take the aforementioned steps. Thus, they need to provide situation patterns describing appropriate usage situations, a service description containing information about the service main purpose, advantages, costs and more. Such information needs to be created and provided by the service providers in a standardized form.

That way, service and product alternatives with assisting information for building implementation intentions can be provided to the user, fitting to his current context. The alternatives can be ranked by the client by different factors (e.g. the service situation fitness, the utility of the alternatives or personal preferences) and visualize the data. The user can give implicit or explicit feedback, which can be used by the client to optimize the proposed alternatives. An example following the proposed approach will be given in the next section.

6 Example: "Lend me"- plugin for online shopping webpages

In this section a simplified example of the proposed system is introduced to shine a light on possible applications and for clarification of the introduced conceptual architecture (see Fig. 3).

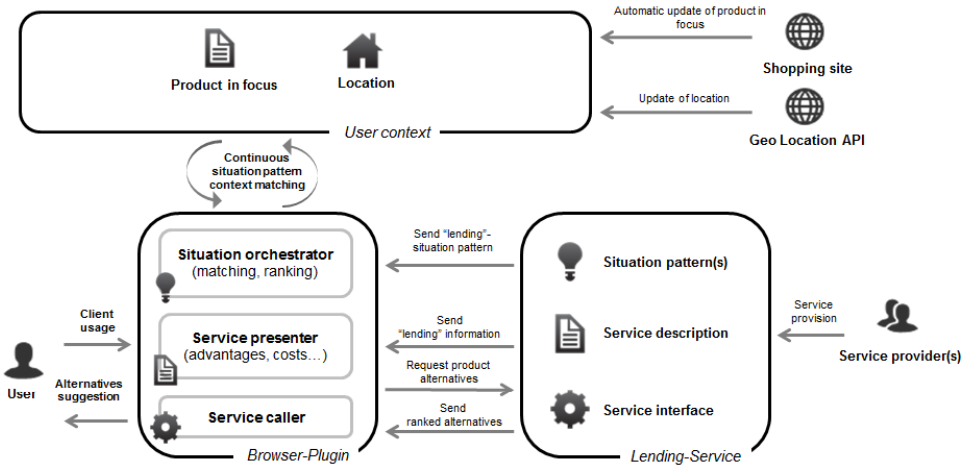


Fig. 3: Exemplary application of SALT

To reduce complexity, the presented example involves only one particular lending service. If several services were involved the question of how the services compete among each other regarding the ranking of appropriate offers would arise. Although this question is not answered yet, it's not in the focus of this publication, but an open issue to be addressed in our further research. The goal of the lending service is to provide the user with information about people who are willing to lend a product, which the user is intending to buy on an online shopping site. The client is therefore realized as a browser plugin, which annotates specific shopping websites with alternatives relevant for the user when the same is currently in an online shopping situation.

The task of the situation orchestrator is to recognize such a situation. Therefore the client-side browser plugin needs at least information about the product in focus, i.e. the current website, a unique product identifier (in this case the EAN, the European Article Number), the product class and the product price. On the other side, the lending service needs to provide situation patterns that can be supported by the service to find an appropriate matching. A simple matching solution based on semantic technologies in particular a rule-based system, is presented in the following. The JENA framework[3] and its included inference engine with a forward chaining RETE [Fo82] rule engine could be

[3] http://jena.apache.org

used to solve the matching. The following example is formalized in RDF[4] using the N3[5] notation.

```
Semantic context information (excerpt):
:c rdf:type :Context ;
  :currentWebsite :shoppingParadise .
:shoppingParadise rdf:type :Website ;
  :websiteType :shopping ;
  :websiteURL http://shoppingParadise.com .
:c :productInFocus :product .
:product rdf:type :Product ;
  "ABC Inc. cordless screwdriver" ;
  :class :consumerProduct ;
  :price 79.00 ;
  :currency :euro ;
  :ean 0123456789012 .

Semantic situation rules of lending service (simplified):
(?c rdf:type :Context), (?c :currentWebsite ?w),
(?w rdf:type :Website), (?w :websiteType :shopping),
(?p rdf:type :Product), (?c :productInFocus ?p),
(?p :class :consumerProduct), (?p :ean ?ean)
  -> service fits to context
```

The context, as described above, contains the information that the user is currently visiting a shopping website and thinking about buying a cordless screwdriver. The rules provided by the lending service are matching because the current context is checked for the user being on a shopping website and visiting a consumer product, which also provides the EAN, which is later on required to provide alternatives.

After a match between the situation pattern and the context is identified the service description of the service is presented via the service presenter, which can appear as an annotation on the shopping website.

Simultaneously the client requests the lending service to construct specific alternatives and embeds the context in this request. The lending service searches the database of the lending system provider for the product EAN and product class and extracts the people who are lending this product.

The next step is the ranking of this set of people, which can be done generally by computing a utility score for each person. A simple way to do this is using a weighted aggregation of different feature scores. For this system three feature scores are proposed, which follow the three dimensions of the sustainability triangle:

Environmental score: The spatial distance between the user and lender divided by the maximal distance between the user and a lender in the set of possible lenders.

[4] http://www.w3.org/standards/techs/rdf#w3c_all
[5] http://www.w3.org/TeamSubmission/n3/

Social score: The rating of this person for this kind of transactions divided by the maximal rating.

Economic score: The cost savings divided by the price of buying this product.

After ranking the alternatives in such a way, the lending service returns this ranking to the client. Finally the client presents the alternatives to the user, who can click on the different alternatives and gets forwarded directly to the lending offer.

7 Summary, challenges and future research

The goal of the paper was to introduce a new type of application in the field of Sustainable by ICT, namely the sustainable service and product alternatives system, the SALT system. Grounded on a psychological and social model of the transformation from intentions to behavior, a conceptual architecture was proposed and illustrated with the example of a browser plugin for the presentation of alternatives to the purchasing of goods in a web-based scenario.

As this paper is a work-in-progress paper, there are open issues, which will be addressed in subsequent work. Among others, a way to rank different services and a feedback mechanism to personalize the presented alternatives has to be developed, as well as the further development of the initial proposals in this paper. Furthermore topics like privacy issues, the usability and the way of presenting the alternatives have to be considered.

References

[Ab15] Abdullah, R. et al.: A Systematic Literature Review of Green Software Development in Collaborative Knowledge Management Environment. In International Journal of Advanced Computer Technology (IJACT), 2015, 9; S. 136.

[AD07] Auger, P.; Devinney, T. M.: Do What Consumers Say Matter? The Misalignment of Preferences with Unconstrained Ethical Intentions. In Journal of Business Ethics, 2007, 76; S. 361–383.

[Aj91] Ajzen, I.: The theory of planned behavior. In Organizational Behavior and Human Decision Processes, 1991, 50; S. 179–211.

[AM86] Ajzen, I.; Madden, T. J.: Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. In Journal of Experimental Social Psychology, 1986, 22; S. 453–474.

[ANC12] Agarwal, S.; Nath, A.; Chowdhury, D.: Sustainable Approaches and Good Practices in Green Software Engineering. In International Journal of Research and Reviews in Computer Science, 2012, 3; S. 1425–1428.

[Be75] Belk, R. W.: Situational Variables and Consumer Behavior. In Journal of Consumer Research, 1975, 2; p. 157.

[CNW10] Carrington, M. J.; Neville, B. A.; Whitwell, G. J.: Why Ethical Consumers Don't Walk Their Talk: Towards a Framework for Understanding the Gap Between the Ethical Purchase Intentions and Actual Buying Behaviour of Ethically Minded Consumers. In Journal of Business Ethics, 2010, 97; p. 139–158.

[DSB10] DiSalvo, C.; Sengers, P.; Brynjarsdóttir, H.: Mapping the landscape of sustainable HCI. In (Mynatt, E. et al. Hrsg.): CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

[Fe02] Federal Government: Perspectives for Germany. Our strategy for Sustainable Development, 2002.

[Fe11] Fennis, B. M. et al.: Bridging the intention–behavior gap: Inducing implementation intentions through persuasive appeals. In Journal of Consumer Psychology, 2011, 21;

[Fo02] Fogg, B. J.: Persuasive technology. In Ubiquity, 2002.

[Fo82] Forgy, C. L.: Rete: A fast algorithm for the many pattern/many object pattern match problem. In Artificial Intelligence, 1982, 19; p. 17–37.

[Fu05] Futtera, S. C. L.: The rules of the game: The principal of Climate Change Communication, London, 2005.

[GKB09] Gustafsson, A.; Katzeff, C.; Bang, M.: Evaluation of a pervasive game for domestic energy engagement among teenagers. In Computers in Entertainment, 2009.

[Go99] Gollwitzer, P. M.: Implementation intentions: Strong effects of simple plans. In American Psychologist, 1999, 54; p. 493–503.

[HH15] Huber, M. Z.; Hilty, L. M.: Gamification and Sustainable Consumption: Overcoming the Limitations of Persuasive Technologies. In (Hilty, L. M.; Aebischer, B. Hrsg.): ICT Innovations for Sustainability. Springer International Publishing, Cham, 2015; p. 367–385.

[Hu10] Hutchison, D. et al. Hrsg.: Persuasive Technology. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[Ma12] Maass, W. et al.: Applying situation-service fit to physical environments enhanced by ubiquitous information systems. In ECIS 2012 Proceedings, 2012.

[Pe13] Penzenstadler, B.: Towards a Definition of Sustainability in and for Software Engineering: Proceedings of the 28th Annual ACM Symposium on Applied Computing, 2013; p. 1183–1185.

[Pe14] Penzenstadler, B. et al.: Systematic mapping study on software engineering for sustainability (SE4S). In (Shepperd, M.; Hall, T.; Myrtveit, I. Hrsg.): EASE '14. Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, 2014; p. 1–14.

[Sh02] Sheeran, P.: Intention—Behavior Relations: A Conceptual and Empirical Review. In European Review of Social Psychology, 2002, 12; p. 1–36.

[We15] WeAct Challenge. http://www.weact.ch/weact-challenge/, retrieved on 03.05.2015.

Douglas Cunnigham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Konzeption und Entwicklung eines Genehmigungsmanagements für KMU unterstützt durch die Augmented Reality

Patrick Piltz[1], Volker Wohlgemuth[2]

Abstract: Die Augmented Reality (AR) ist ein spannender Bereich für die Erkennung von Objekten, ohne das weitere Maßnahmen des Anwenders benötigt werden. Dieses Verfahren ist nicht wirklich neu, da schon in den 1990er Jahren viele Artikel zu diesem Thema veröffentlicht wurden. Jedoch erfährt die AR erst seit einigen Jahren ein steigendes Interesse. Viele Anwendungen insbesondere mobile Applikationen verwenden AR-Ansätze. Für die betriebliche Umweltinformatik spielt die AR bisher jedoch kaum eine Rolle. In diesem Beitrag wird deshalb die Anwendbarkeit von AR-Ansätzen im Kontext der betrieblichen Umweltinformatik analysiert und eine Anwendung im Anwendungsfeld des Genehmigungsmanagements von Anlagen nach dem Bundes-Immissionsschutzgesetz (BImSchG) prototypisch umgesetzt.

Keywords: Augmented Reality, künstliche Marker, natürliche Marker, Genehmigungsmanagement

1 Einleitung

Die Augmented Reality (AR) findet auf dem Gebiet der betrieblichen Umweltinformatik bisher wenig Beachtung. Dabei bietet der Einsatz von AR-Methoden einige Potentiale [Me11], S. 24:

- Rechnergestützte Erweiterung der menschlichen Wahrnehmung,
- Visualisierung von Informationen,
- Minimierung der Zeit zur Informationsbeschaffung (Time-to-Content) und
- Kombination von haptischem und digitalem Erlebnis (hap.dig).

Es bietet sich daher an, dass sich die betriebliche Umweltinformatik diese Potentiale zu Nutze macht. Zumindest der dritte Punkt die Informationsbeschaffung spielt oftmals auch im Kontext der betrieblichen Umweltinformatik eine wichtige Rolle, z.B. für den Bereich des Genehmigungsmanagements von Anlagen.

In diesem Beitrag wird deshalb die AR im Kontext des Genehmigungsmanagements erprobt. Dabei werden verschiedene Trackingverfahren angewandt und mit einer her-

[1] Hochschule für Technik und Wirtschaft Berlin, Fachbereich 2: Ingenieurwissenschaften – Technik und Leben, Wilhelminenhofstr. 75a, 12459 Berlin, patrick.piltz@hotmail.de
[2] Hochschule für Technik und Wirtschaft Berlin, Fachbereich 2: : Ingenieurwissenschaften – Technik und Leben, Wilhelminenhofstr. 75a, 12459 Berlin, volker.wohlgemuth@htw-berlin.de

kömmlichen Suchmethode nämlich Listen mit Filterfunktion verglichen.

Für diesen Zweck wurde ein Softwareprototyp entwickelt. Dieser besteht aus einer Server-Komponente für die Datenhaltung einem Desktop-Client, welcher zur Verwaltung und Analyse von Daten und zur Planung von Prüfungen dient, einem mobilen Client für die Durchführung von Prüfungen vor Ort an einer Anlage und einem zweiten mobilen Client zur Erprobung verschiedener Trackingverfahren der AR. Das gesamte System soll zum einen den Anwender bei der Verwaltung der Genehmigungen der Anlagen unterstützen und zum anderen helfen das Potential der AR für den Bereich der betrieblichen Umweltinformatik abzuschätzen.

Für die Implementierung des Genehmigungsmanagements wird auf die Ergebnisse des OpenResKit-Projektes der Hochschule für Technik und Wirtschaft in Berlin zurückgegriffen, welches ein Open-Source-Framework für die Implementierung von Client/Server-Systemen bereitstellt.

2 Grundlagen

2.1 Bundes-Immissionsschutzgesetz

Ein Ziel des Umweltschutzes ist die Reduzierung von Emissionen. Hierbei existieren für jedes Kompartiment eigene Gesetze. Diese sind [AS09], S. 1:

- das Bundes-Bodenschutzgesetz,
- das Bundes-Immissionsschutzgesetz,
- das Bundes-Naturschutzgesetz und
- das Wasserhaushaltsgesetz

In diesem Beitrag wird vorrangig das Bundes-Immissionsschutzgesetz (BImSchG) behandelt. Auch die Software wurde in dieser Hinsicht konzipiert, lässt sich aber dank des zugrunde liegenden Frameworks leicht erweitern.

Für die Anwendung des BImSchG ist zuerst einmal wichtig was im Sinne des Gesetzes eine Anlage ausmacht. Nach § 3 Abs. 5 BImSchG muss für eine Einstufung als Anlage eines der folgenden Merkmale erfüllt sein. Bei einer Anlage handelt es sich um:

- eine betriebsfeste und sonstige ortsfeste Einrichtung,
- Maschinen, Geräte und sonstige ortsveränderliche technische Einrichtungen oder
- Grundstücke, auf denen Stoffe gelagert oder abgelagert oder emissionsträchtige Arbeiten durchgeführt werden, ausgenommen öffentliche Verkehrswege.

Besonders wichtig für das Genehmigungsmanagement sind folgende Paragraphen des BImSchG [AS09], S. 74:

- §§ 1 - 3 Zweck, Geltungsbereich und Begriffsdefinition
- §§ 4 - 21 Genehmigungsbedürftige Anlagen
- §§ 22 -25 Nicht genehmigungsbedürftige Anlagen

In diesem Beitrag wird auf die §§ 4 - 21 eingegangen, weil genau diese Anlagen mit der Anwendung verwaltet werden sollen. Nach § 4 Abs. 1 Satz 1 BImSchG bedürfen Anlagen einer Genehmigung, „die auf Grund ihrer Beschaffenheit oder ihres Betriebs in besonderem Maße geeignet sind, schädliche Umwelteinwirkungen hervorzurufen oder in anderer Weise die Allgemeinheit oder Nachbarschaft zu gefährden, erheblich zu benachteiligen oder erheblich zu belästigen“ [DT12] § 4 Abs. 1 BImSchG. Diese Anlagen müssen ein Genehmigungsverfahren durchlaufen. Anhand gewisser Kriterien wird das vereinfachte oder das förmliche Verfahren durchgeführt.

2.2 Augmented Reality

Das Realitäts-Virtualitäts-Kontinuum entwickelt von Milgram et al. bietet einen guten Einstieg in die AR. Darin ist zu erkennen, dass die Mixed Reality aus der Augmented Reality und der Augmented Virtuality besteht. Die Mixed Reality reicht von der realen Umgebung bis zur virtuellen Umgebung. Beide stellen die Extrempunkte der Mixed Reality dar. Die AR ist dort, wo der Anteil realer Objekte höher ist als der Anteil virtueller Objekte. Es wird demnach die reale Umgebung um virtuelle Objekte erweitert. Die Augmented Virtuality ist analog zu betrachten nur umgekehrt [Mi94] S. 283.

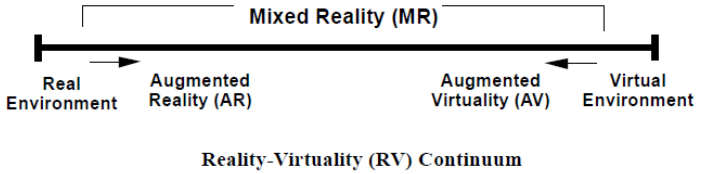


Abb. 1: Realitäts-Virtualitäts-Kontinuum nach Milgram, et al. [Mi94], S. 283

Nach Azuma können der Augmented Reality drei Charakteristika zugeschrieben werden [Az97], S. 2 f.:

- Kombination von virtueller Realität und realer Umwelt
- Interaktion in Echtzeit
- Dreidimensionaler Bezug von Objekten

Die in diesem Beitrag vorgestellten Lösungsansätze für AR entsprechen nicht den von Azuma aufgestellten Charakteristika, da hier kein dreidimensionaler Bezug besteht sondern lediglich ein zweidimensionaler Bezug. Diese Form der AR wird als Augmented Reality im weiteren Sinne bezeichnet. [Me11] S. 11

2.3 Trackingverfahren

Für die Identifizierung von Objekten mit Hilfe von AR-Methoden wird in zwei Vorgehensweisen unterschieden. Da gibt es zum einen das nichtvisuelle Tracking. Dieses Verfahren verzichtet vollständig auf die Bildverarbeitung. Vielmehr greift dieses Verfahren auf Komponenten heutiger mobiler Endgeräte zurück, wie zum Beispiel in [Ro01] aufgeführt:

- Ultraschall
- GPS
- Optische Sensoren (z. B. Infrarot)
- Beschleunigungssensor
- Kompass

Anhand dieser Daten können die Position des Anwenders und seine Blickrichtung ermittelt werden. So können dem Anwender die entsprechenden Daten angezeigt werden, die dieser gerade benötigt.

Die zweite Vorgehensweise ist das visuelle Tracking [Me11] S. 27. In diesem Verfahren wird hauptsächlich mit den Daten der Kamera des mobilen Endgerätes gearbeitet. Auch hier gibt es zwei verschiedene Ansätze des Trackings. Zum einen das Tracking mit künstlichen Marker und zum anderen das Tracking ohne Marker.

Ersteres benötigt Bilder, die in der realen Umgebung platziert werden. Hierbei kann es sich zum Beispiel um QR-Codes handeln. So können vor Ort die Daten eines Objektes ermittelt werden. Dies ist ein sehr zuverlässiges Verfahren, um die gewünschten Daten zu erhalten [Ow02] S. 1. Jedoch bringt diese Methode auch einen sehr großen Aufwand mit sich, denn die Bilder müssen platziert und gewartet werden. Deshalb eignet sich diese Methode eher für einen kleinen Einsatzbereich. [Kl06] S. 4

Das visuelle Tracking ohne Marker verzichtet auf diese künstlichen Marker. Vielmehr greift dieser Ansatz auf Marker zurück, die bereits vorhanden sind nämlich natürliche Marker wie Ecken und Kanten. So können Objekte zum Beispiel anhand von Bildern der Kamera oder anhand von CAD-Modellen (Computer Aided Design) [Bl06] S. 56 erkannt werden. Ein Verfahren, das komplett ohne künstliche und natürliche Marker und ohne Modelle auskommt ist das PTAM-Verfahren (Parallel Tracking and Mapping). Mit Hilfe der Bilder der Kamera wird eine visuelle Karte erzeugt. Sobald diese grob vorhanden ist, wird diese Karte mit virtuellen Objekten gefüllt und den realen Objekten zugeordnet. So kann auch eine unbekannte Umgebung in relativ kurzer Zeit mit mobilen Geräten erkannt werden. [KM10] S. 1

2.4 Einsatzmöglichkeiten im Kontext der betrieblichen Umweltinformatik

Eine Möglichkeit für die Anwendung von AR-Methoden im Kontext der betrieblichen

Umweltinformatik bietet das Stoffstrommanagement. Der Anwender zeigt mit der Kamera auf eine Anlage. Diese Anlage liefert nun alle Daten zum In- und Output. Mit Schiebereglern können dann die einzelnen In- und Outputs verändert werden. So können anschließend die Änderungen angezeigt werden, wie z.B. der veränderten Energieverbrauch mitsamt den hieraus resultierenden Emissionen. Sind die neuen Daten besser als die Vorherigen und wirken sich diese nicht negativ auf die Produktion aus, können Alternativen vor Ort visualisiert und mit den betroffenen Mitarbeitern in Bezug auf ihre Praktikabilität diskutiert werden.

Ein weiteres Anwendungsfeld der betrieblichen Umweltinformatik ist die IT-Unterstützung von Gefährdungsbeurteilungen von Arbeitsplätzen. Nach § 5 Arbeitsschutzgesetz ist jeder Arbeitgeber verpflichtet, die Arbeitsplätze nach potentiellen Gefährdungen zu untersuchen. Beispielsweise kann der Anwender nach einer bereits durchgeführten Gefährdungsbeurteilung die Maßnahmen umsetzen, die aus dieser resultierten, damit die Gefährdungen vermieden oder zumindest vermindert werden. In dem Kamerabild sieht der Anwender alle analysierten Gefährdungen des Arbeitsplatzes, indem die möglichen Gefahrenstellen des Arbeitsplatzes im Kamerabild durch ein Ausrufezeichen markiert sind. Diese werden zusätzlich farblich unterschiedlich in Bezug auf ihren Schweregrad visualisiert. Durch einen Klick auf ein Ausrufezeichen erhält der Anwender Informationen, wie diese Gefahrenpotentiale abgestellt werden könnten. So weiß der Anwender auf einen Blick, um welche Gefährdungen er sich zuerst kümmern muss und welche gegebenenfalls später bearbeitet werden können.

2.5 OpenResKit-Framework

Das OpenResKit-Framework entstand aus dem OpenResKit-Projekt, das an der Hochschule für Technik und Wirtschaft in Berlin (HTW Berlin) durchgeführt worden ist. Ziel dieses Projektes war es, kleinen und mittleren Unternehmen einen Open-Source-Softwarebaukasten an die Hand zu geben, den diese selber an die eigenen Bedürfnisse anpassen können und so die betrieblichen Prozesse des Unternehmens transparenter in Bezug auf stoffliche und energetische Aspekte zu machen.

Im Rahmen des OpenResKit-Projektes sind schlanke, problemspezifische und einfache Softwarebausteine entstanden, die die Betriebe unterstützen, eine strukturierte Datengrundlage für die genannten Zwecke aufzubauen. Hierfür wurden Desktop-Anwendungen für die Verwaltung und Analyse der Daten entwickelt sowie mobile Anwendungen für die Datenerhebung vor Ort. Die Realisierung einer Software egal ob mobil oder stationär ist nicht an eine Programmiersprache gebunden, da der Hub durch WCF-Services (Windows Communication Foundation) bekannte Schnittstellen bereitstellt, die mit verschiedensten Programmiersprachen angesteuert werden können.

Zentraler Bestandteil dieses Frameworks ist der Hub, welcher mit den Clients kommuniziert. Dies kann in Form von WCF-Services wie zum Beispiel der Austausch von Daten oder zur Erzeugung von Beispieldaten erfolgen. Hierbei existieren zwei Arten der Kommunikation zwischen Client und Hub, nämlich eine asynchrone und eine synchrone

Kommunikation. Die synchrone Kommunikation wird für die Desktop-Anwendungen genutzt, da davon ausgegangen werden kann, dass ein ständiger Internetzugang existiert. Im Gegensatz dazu wird die asynchrone Kommunikation für die mobilen Anwendungen genutzt, da es hier wahrscheinlich ist, dass nicht zu jeder Zeit ein Internetzugriff vorliegt.

Somit dient der Hub als Datenquelle und -senke. Ein positiver Aspekt dieses Frameworks ist, dass Medienbrüche vermieden werden, da alle Daten egal ob diese von der Desktop-Anwendung oder der mobilen Anwendung stammen zentral im Hub persistiert werden. Zusätzlich kann die Domäne des Hubs leicht kombiniert und durch neue eigens entwickelten Plug-Ins erweitert werden, um so neue Prozesse abbilden zu können [Wo14].

Die nachfolgende Abbildung verdeutlicht die Architektur des OpenResKit-Frameworks.

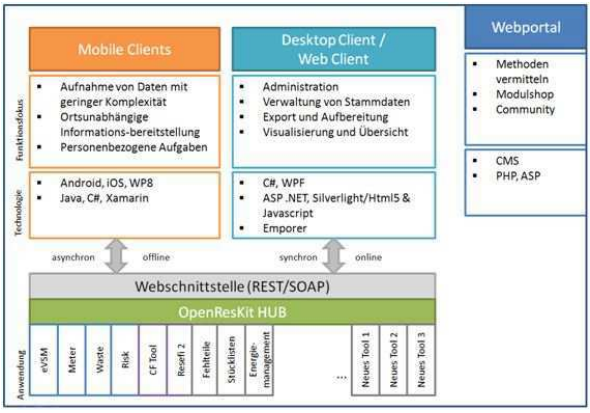


Abb. 2: Software-Architektur des OpenResKit-Frameworks [Wo14]

3 Konzeption

Für die Implementierung des Genehmigungsmanagements wurde sich stark an der Vision des OpenResKit-Projektes orientiert. Das heißt, neben dem Hub existieren diverse schmale und funktionale Clients. Jeder Client berücksichtigt dabei einen dedizierten Bearbeitungsschritt im Rahmen des Genehmigungsmanagements.

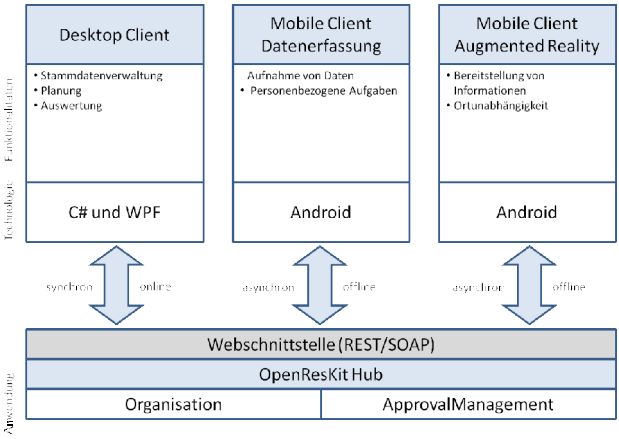


Abb. 3: Architektur des Softwaresystems zur Unterstützung des Genehmigungsmanagements

In der Abbildung 2 ist zu erkennen, dass der Hub eine zentrale Rolle innerhalb der Softwarearchitektur inne hat. Für seine Implementierung werden bereits vorhandene Bibliotheken des OpenResKit-Projekts verwendet. Dieser tauscht Daten mit drei verschiedenen Clients aus. Zum einem der Desktop-Client der für die Verwaltung von Daten, Planung von Prüfungen sowie deren Auswertung zuständig ist. Zum anderen der erste mobile Client der für die Erfassung von Daten vor Ort an der Anlage zuständig ist. Zuletzt existiert ein zweiter mobiler Client, der von den Anlagenfahrern genutzt werden kann, um schnell auf die Informationen der Anlage zugreifen zu können. Beide mobile Clients kommunizieren zum Hub asynchron weil kein ständiger Internetzugriff besteht.

4 Realisierung

4.1 Hub

Für den Hub wurde auf einen bestehenden Softwarebaustein des OpenResKit-Projektes zurückgegriffen. Hierbei handelt es sich um den OpenResKit-Hub. Dieser wurde um eine eigene Domäne für das Genehmigungsmanagement erweitert. Für diese Erweiterung wurden ganz normale Klassen erzeugt. Damit deren Daten in der Datenbank gespeichert werden können, müssen diese mit Hilfe des Entity Frameworks konfiguriert werden. Auf diese Weise werden automatisch für die Klassen mit ihren Eigenschaften entsprechende Tabellen in der Datenbank erzeugt.

4.2 Desktop-Client

Der Desktop-Client wurde genau wie der Hub in der Programmiersprache C# von .NET entwickelt. Als Auszeichnungssprache wurde WPF (Windows Presentation Foundation)

wegen der moderneren Oberfläche gewählt. Der Desktop-Client erhält die Daten vom Hub über einen WCF Data Service (Odata) und lädt diese in ein lokales Repository. Diese Daten können anschließend erweitert, geändert oder gelöscht werden.

Wie bereits erwähnt dient der Desktop-Client hauptsächlich der Verwaltung der Stammdaten wie Anlagen und Genehmigungen. Hierfür wird dem Anwender ein Formular bereitgestellt, in dem die benötigten Daten eingetragen werden können.

Neben der Verwaltung der Stammdaten ist die Auswertung der Prüfungen eine zentrale Funktion des Desktop-Clients. Für einen schnellen Überblick der Prüfungen werden diese in auffällige und unauffällige Prüfungen gruppiert. Bei den auffälligen Prüfungen können dann die geprüften Nebenbestimmungen mit ihren Ergebnissen betrachtet und dafür Maßnahmen generiert werden.

4.3 Mobiler Client zur Datenerfassung

Der mobile Client für die Datenerfassung dient der Durchführung von Prüfungen vor Ort an der Anlage. Ziel dieses mobilen Clients ist die papier- und stiftlose Erfassung von Daten. Diese Daten können an den Hub geschickt werden, wodurch ein Medienbruch vermieden wird.

Hierfür werden dem Anwender zuerst alle geplanten Prüfungen angezeigt. Dieser wählt seine Prüfung aus und kann die Nebenbestimmungen auf ihre Einhaltung überprüfen. Bei einer Nichteinhaltung kann der Anwender Maßnahmen hinterlegen, die ausgeführt werden müssen, damit die Nebenbestimmung wieder eingehalten wird.

4.4 Mobiler Client zur Bereitstellung von Daten mittels Augmented Reality-Methoden

Der mobile Client für die Bereitstellung von Informationen dient in erster Linie der Erprobung von AR-Methoden im Kontext der betrieblichen Umweltinformatik. Diese Applikation bietet drei Möglichkeiten für das Suchen nach einer Anlage. Zum einen die zwei genannten Trackingansätze aus Abschnitt 2.3 - einmal mit künstlichen Marker, einmal ohne und das Suchen aus einer Liste. Auf die Suche aus der Liste wird hier nicht weiter eingegangen.

Die Suche mit künstlichen Markern ist heutzutage relativ einfach umzusetzen, besonders wenn die Applikation auf Basis des Androidsystems entwickelt wird. Hier existieren etliche open-source Softwareprodukte, die die Erkennung von QR-Codes und Barcodes mittels Kamera unterstützen. Solch eine Software (zxing) [zx09] wurde für diese Applikation genutzt. Sie ist robust und liefert zuverlässige Ergebnisse.

Für den zweiten Lösungsansatz sollte es nicht Bedingung sein, vorher QR-Codes oder Barcodes zu platzieren. Eher sollten die Anlagen natürlich erkannt werden. Problem ist, dass an einer Anlage nichts wirklich heraussticht. Erschwert wird die Problematik, wenn

von einem Anlagentyp mehrere Exemplare vorhanden sind. Diese können dann nicht mehr eindeutig über die Bilder der Kamera identifiziert werden. Daher wird sich bei diesem Ansatz auf eine Annahme gestützt. Es wird angenommen, dass auf jeder Anlage eine Anlagennummer aufgedruckt ist. Diese dient dann als natürlicher Marker. Über die Bilder der Kamera läuft eine Texterkennungssoftware. Wenn ein Text erkannt wird, wird dieser in Form eines Strings für die Filterung der Anlagen genutzt. Der Suchtreffer wird anschließend dem Anwender dargestellt.

5 Bewertung und Ausblick

Beide AR-Lösungen, die Identifizierung mit QR-Code und die Identifizierung nach den Anlagennummern, liefern schnell und verlässlich die benötigten Informationen zu der erkannten Anlage. Hinzuzufügen ist, dass der AR-Lösungsansatz mit dem QR-Code robuster und zuverlässiger ist. Jedoch hat auch der AR-Lösungsansatz, der Identifizierung nach der Anlagennummer, seinen Reiz, da das Platzieren und Warten der Bilder (z. B. QR-Codes) für die Augmented Reality nicht erforderlich sind. Es steht außer Frage, dass in diesem AR-Lösungsansatz noch Optimierungen und Erweiterungen von Nöten sind. Dennoch ist dies eine sehr gute Alternative zu den QR-Codes und dem Suchen aus einer Liste vorzuziehen. Hier ist schon bei der Identifizierung des Objektes eine erhebliche Zeitersparnis festzustellen. Die Suche aus einer Liste ist oftmals ein mühseliges und zeitaufwändiges Unterfangen insbesondere auf einem mobilen Endgerät mit kleinem Display.

Es sei angemerkt, dass die Erfolgswahrscheinlichkeit des gewählten AR-Ansatzes zur Erkennung von Objekten nicht bei 100 Prozent liegt. Grund hierfür ist, dass die Texterkennungssoftware, die über die Bilder der Kamera läuft, Zeichen erkennt, die nicht vorhanden sind oder vorhandene Zeichen falsch interpretiert. So kann es passieren, dass falsche Objekte dargestellt werden. Eine Möglichkeit dieses Problem zu umgehen ist, dass eine Art Rangliste mit Objekten erstellt wird, deren Anlagennummer dem erhaltenen String ähnlich ist. Die Objekte mit der höchsten Übereinstimmung werden dem Anwender präsentiert. Dieser braucht dann lediglich das richtige Objekt auszuwählen.

Die Verwendung von AR-Ansätzen bietet auch im Kontext der betrieblichen Umweltinformatik interessante Ansätze. Besonders in Fällen in denen Daten in regelmäßigen Abständen erfasst werden müssen, scheint ihr Einsatz sinnvoll, da eine schnelle Identifizierung der Objekte hier enorme Zeitersparnisse erzielen können. Der in diesem Beitrag vorgestellte Softwareansatz beschreibt, wie AR-Methoden im Kontext der betrieblichen Umweltinformatik sinnvoll eingesetzt werden können. Auch wenn die Ergebnisse noch nicht voll zufriedenstellend sind, sind hier die Potentiale klar erkennbar: eine schnelle und einfache Bereitstellung von Informationen und eine leichte Interaktion mit den Objekten (siehe Beispiel zwei in Kapitel 2.4 Einsatzmöglichkeiten im Kontext der betrieblichen Umweltinformatik).

Literaturverzeichnis

[AS09] ALPMANN Schmidt: Umweltrecht Naturschutz, Immissionsschutz, Abfallrecht, Bodenschutz, Wasserrecht. ALPMANN Schmidt, 2009.

[Az97] Azuma, Ronald T.: A Survey of Augmented Reality. Teleoperators and Virtual Environments 6, 4, 355 - 385, 1997.

[Bl06] Bleser, Gabriele; Wuest, Harald; Stricker, Didier: Online camera pose estima tion in partially known and dynamic scenes. 2006.

[DT12] Deutscher Taschenbuch Verlag: Umweltrecht. Deutscher Taschenbuch Verlag, 2012.

[Kl06] Klein, Georg: Visual Tracking for Augmented Reality. http://www.robots.ox.ac.uk/~gk/publications/Klein2006Thesis.pdf, Stand: 10.04.2015.

[KM10] Klein, Georg; Murray, David: Parallel Tracking and Mapping for Small AR Workspaces. http://www.robots.ox.ac.uk/~lav/Papers/klein_murray_ismar2007/klein_murray_ismar2007.pdf, Stand: 10.04.2015.

[Me11] Mehler-Bicher, Anett; Reiß, Michael; Steiger, Lothar: Augmented Reality. Oldenbourg Wissenschaftsverlag GmbH, 2011.

[Mi94] Milgram, Paul; Takemura, Haruo; Utsumi, Akira; Kishino, Fumio: Augmented Reality: A class of displays on the reality-virtuality continuum. http://web.cs.wpi.edu/~gogo/hive/papers/Milgram_Takemura_SPIE_1994.pdf, Stand: 10.04.2015.

[Ow02] Owen, Charles B.; Xiao, Fan; Middlin, Paul: What is the best fiducial?. ftp://debian.cse.msu.edu/CSE/pub/multimedia/charles-owen-art02.pdf, Stand: 10.04.2015.

[Ro01] Rolland, Jannick P.; Baillot, Yohan; Goon, Alexei A.: A SURVEY OF TRACKING TECHNOLOGY FOR VIRTUAL ENVIRONMENTS. http://www.creol.ucf.edu/Research/Publications/1522.PDF, Stand: 10.04.2015.

[Wo14] Wohlgemuth, Volker: OpenResKit. http://openreskit.htw-berlin.de/technische-vision, Stand: 10.04.2015.

[Zx09] zxing: zxing. https://github.com/zxing/zxing, Stand: 10.04.2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Ist ein Perspektivenwechsel der Umweltinformatik durch ein mögliches Verschwinden der Grenze zwischen physischem Objekt und Software notwendig?

Hans-Knud Arndt[1]

Abstract: In der (Umwelt-)Informatik ist die Bestimmung der Grenze zwischen physischem Objekt und Software eine zentrale Problemstellung. Der Trend zur Auflösung dieser starren Grenzen wie z.B. bei dem Produkt der Apple Watch führt zu Fragestellungen an die Umweltinformatik. Eine Konsequenz aus dem Verschwinden der Grenze zwischen physischem Objekt und Software liegt in der Konvergenz der jeweiligen Gestaltungsprinzipien und damit in einem Perspektivwechsel der Gestaltungsprozesse in der (Umwelt-)Informatik.

1 Innovationen für Benutzerschnittstellen der Informations- und Kommunikationstechnik am Beispiel des Unternehmens Apple Inc.

Das US-amerikanische Unternehmen Apple Inc. zeichnet sich dadurch aus, dass es ihm gelingt neue Gebiete der Informations- und Kommunikationstechnik mit einer hohen Marktdurchdringung zu erschließen, nachdem Konkurrenzunternehmen dies vergeblich versucht hatten und scheiterten [Mos14]. Dies gilt insbesondere für das Gebiet der Benutzerschnittstellen:

- **1983/1984:** Einführung der grafischen Benutzeroberfläche und der Maus in den mit den Computern Lisa (1983) und Macintosh (1984) [OTW13].
- **2004:** Einführung des tragbaren digitalen Medienabspielgeräts iPod mini mit einem berührungsempfindlichen Scrollrad („Touch Wheel“) in Verbindung mit Tasten in einer Oberfläche, dem „Click Wheel“ [App04].
- **2007:** Einführung von Smartphones durch das „iPhone“ mit intelligentem „Multi-Touch-Bildschirm“, wobei die Bedienung in Verbindung mit einem Schalter („Home Button“) praktisch vollständig durch Gesten auf dem berührungsempfindlichen Bildschirm erfolgt [App09].

Das neue Produkt „Apple Watch“, welches erstmals offiziell auf dem Apple Event am 9. September 2014 in Cupertino vorgestellt wurde, soll wiederum eine "völlig neue Benutzeroberfläche, die speziell für ein Gerät, das entworfen ist, um getragen werden“

[1] Otto-von-Guericke Universität Magdeburg, Fakultät für Informatik, Arbeitsgruppe Wirtschaftsinformatik – Managementinformationssysteme, hans-knud.arndt@iti.cs.uni-magdeburg.de

[App14], auf dem Markt etablieren.

2 Die Apple Watch und das neue Bedienkonzept

Das neue Produkt „Apple Watch“ (siehe Abbildung 1) ist entsprechend der Definition von Apple ein „unglaublich präziser Zeitmesser, ein persönliches und direktes Kommunikationsgerät und ein bahnbrechender Begleiter für Gesundheit und Fitness“ [App15a]. Mit der Apple Watch werden folgende neue Bedienkonzepte eingeführt [App15a]:

- **Digitale Krone (Digital Crown):** Bei herkömmlichen Uhren stellt die Krone ein „geriffeltes Rädchen, geriffelter Knopf an Armband- oder Taschenuhren zum Aufziehen des Uhrwerks oder zum Stellen der Zeiger“ [Dud13b] dar. Apple setzt im Rahmen des Bedienkonzepts der Apple Watch die „Digitale Krone“ ein, um einen innovativen Weg „zum flüssigen Scrollen, Zoomen und Navigieren ohne das Display zu versperren“ [App15a], bereitzustellen.
- **Retina Display mit Force Touch:** Das Retina Display mit Force Touch ermöglicht es, „nicht nur Berührungen [zu erkennen], sondern auch Druck, was der Benutzeroberfläche eine völlig neue Dimension verleiht. Force Touch nutzt winzige Elektroden rund um das flexible Retina Display, die zwischen leichtem Tippen und stärkerem Drücken unterscheiden können und sofortigen Zugriff auf kontextspezifische Steuerungen erlauben“ [App15b].
- **Taptic Engine:** Die Taptic Engine stellt einen linearen Akuator dar, also ein „Bauelement am Ausgangsteil einer Steuer- oder Regelstrecke, das in Energie- oder Massenströme eingreift und darin als veränderlicher Widerstand wirkt“ [Dud13a]. In der Apple Watch wird die Taptic Engine eingesetzt, um eine haptisches Rückmeldung zu geben. Damit tippt die Apple Watch dem jeweiligen Träger immer auf das Handgelenk, wenn er eine Benachrichtigung oder Mitteilung erhält oder er auf das Display drückt [App15b].



Abbildung 1: Apple Watch
(Quelle: [App15c])

Der Apple Watch dient als Betriebssystem das „Watch OS“, welches auf Apples mobilem Betriebssystem iOS in der Version 8 basiert. iOS stellt die Basis für die portablen

Endgeräte iPhone, iPad und iPod touch von Apple dar. Grundsätzlich vermittelt das Watch OS in Verbindung mit der Digitalen Krone, dem Retina Display mit Force Touch sowie mit der Taptic Engine ein völlig neues Bedienkonzept. Um die Lesbarkeit zu erhöhen, wurde speziell eine neue Schrift für das Watch OS entwickelt. Gleichzeitig erinnert das Bedienkonzept der Apple Watch aber auch optisch an bekannte Gestaltungsmerkmale von iOS 8 [App15b]:

- Einfache, natürliche Gesten wie Zoomen und Streichen.
- Flüssige, benutzerzentrierte Navigation.
- Klar und einfach strukturierte Anordnung der Apps (Anwendungsprogramme).

Das Unternehmen Apple bzw. der Senior Vice President of Design von Apple, Sir Jonathan Ivy, hebt als Besonderheit bei der Vorstellung des Produkts „Apple Watch“ und seinem neuen und innovativen Bedienkonzept hervor, dass die Apple Watch als ein einzigartiges Produkt konzipiert, gestaltet und entwickelt wurde, bei dem sich die **Grenze von physischem Objekt (Hardware) und Software nicht mehr bestimmbar** ist („We conceived, designed, and developed Apple Watch as a completely singular watch. You know you can’t determine a boundary between the physical object and the software.“ [App14]).

In der Informatik ist die Bestimmung der Grenze zwischen Hardware und Software eine zentrale Problemstellung. Die Definition einer solchen Grenze beeinflusste in den letzten Jahrzehnten nicht nur ganz wesentlich die Entwicklung von Informations- und Kommunikationstechnik (IKT) bzw. Informations- und Kommunikationssystemen (IKS), sondern auch die Strukturierung von Forschung und Lehre in der Informatik [PlP13]. Dabei war die Grenze zwischen Hard- und Software zu keiner Zeit als vollkommen fest fixiert anzusehen, d.h., es hat und wird immer Verschiebungen dieser Grenze geben. Apple geht mit der Markteinführung des Produkts „Apple Watch“ aber noch einen Schritt weiter und spricht von einer bis dahin noch nicht gekannten Verschmelzung von Hard- und Software („verschmilzt Apple Watch Hardware und Software wie niemals zuvor“ [App15b]) bzw. von der Auflösung der Grenze von Hard- und Software (siehe oben, „You know you can’t determine a boundary between the physical object and the software.“ [App14]).

3 Verschiebungen der Grenze von physischem Objekt und Software

Unter dem Begriff Hardware werden üblicherweise die physischen IKT-Objekte verstanden, also Objekte die wenigstens vom Grundsatz her als berührbar vorstellbar sind. Ursprünglich wurde der aus dem Englischen stammende Begriff „Hardware“ im Deutschen übersetzt mit Eisenwaren, wie es sie in einem Eisenwarengeschäft zu kaufen gibt. Als Hardware im IKT-Kontext werden z.B. Prozessoren, interne und externe Datenträger, Gehäuse, Tastaturen und Computermäuse angesehen. Dagegen werden als Software

Computerprogramme, gelegentlich auch Daten bezeichnet, also Betrachtungsgegenständer der IKT, die sich nicht berühren lassen. Der Begriff Software ist nicht trennscharf fassbar, vereinzelt wird auch darüber philosophiert, ob Software ohne Hardware überhaupt existieren kann [PlP13].

„Die Suche nach der optimalen Grenze zwischen Hard- und Software ist nach wie vor eine relevante Entscheidung beim Entwurf von Computersystemen, welche – getrieben durch neue Anforderungen und Anwendungen – einer ständigen Veränderung unterliegt“ [PlP13]. Der Trend zur Auflösung einer starren Grenzen zwischen Hardware und Software führt zu einer Reihe neuer Fragestellungen. In technischer Sicht sind das Fragestellungen der Programmierbarkeit, der Modellierung und der Validierung solcher IKT [PlP13].

Aber auch Fragen von der und an die Umweltinformatik sind im Kontext der Verschiebung bzw. Auflösung der Grenze zwischen physischem Objekt und Software von entscheidender Bedeutung. Dies betrifft insbesondere die Gestaltungsprozesse im Rahmen der Umweltinformatik. Die Qualität eines Entwurfs ist der Schlüssel zu mehr Umweltfreundlichkeit und Nachhaltigkeit von Produkten. Dies gilt auch für umweltfreundliche bzw. nachhaltige IKT-Produkte bzw. IT-Ökosysteme [Arn13].

4 Entwurfsprinzipien für die Gestaltung von physischen Objekten und Software

Die Gestaltung von physischen Objekten wird üblicherweise dem Produktdesign bzw. Industriedesign zugeordnet, während die Gestaltung von Software als Frage der Software Ergonomie (Gebrauchstauglichkeit, englisch „Usability“) bzw. als Frage der Ergonomie der Mensch-System-Interaktion angesehen wird. Das gemeinsame Ziel bei der Gestaltung von physischen Objekten als auch von Software ist allgemein formuliert das *gute Design*. Der langjährige und international sehr geschätzte, ehemalige Chefdesigner der Firma Braun AG, Dieter Rams machte sich „zunehmend Sorgen um den Zustand einer Welt, die er als ‚eine undurchschaubare Verwirrung von Formen, Farben und Geräuschen’ empfand (...) [, und stellt sich im] (...) Bewusstsein, dass auch er zu dieser Welt beitrug, (...) eine wichtige Frage: Ist mein Design gutes Design?“[Vit13].

4.1 Prinzipien zur Gestaltung von physischen Objekten

Auf der Suche nach Antworten auf die Frage nach einem guten Design formulierte Dieter Rams auf der Grundlage seiner Erfahrungen auf dem Gebiet des Gestaltens von physischen Objekten zehn Anforderungen an ein gutes Design, die auch als die „Zehn Thesen des Gutes Designs“ bezeichnet werden: „

- **Gutes Design ist innovativ**
  Die Möglichkeiten für Innovation sind noch längst nicht ausgeschöpft. Die technologische Entwicklung bietet immer wieder neue Ausgangspunkte für innovative

Gestaltungskonzepte, die den Gebrauchswert eines Produktes optimieren. Innovatives Design entsteht aber stets im Zusammenhang mit innovativer Technik und ist niemals Selbstzweck. (...)

- **Gutes Design macht ein Produkt brauchbar**
  Man kauft ein Produkt, um es zu benutzen. Es soll bestimmte Funktionen erfüllen – Primärfunktionen ebenso wie ergänzende psychologische und ästhetische Funktionen. Gutes Design optimiert die Brauchbarkeit und lässt alles unberücksichtigt, was nicht diesem Ziel dient oder ihm gar entgegensteht. (...)

- **Gutes Design ist ästhetisch**
  Die ästhetische Qualität eines Produktes ist integraler Aspekt seiner Brauchbarkeit. Denn Geräte, die man täglich benutzt, prägen das persönliche Umfeld und beeinflussen das Wohlbefinden. Schön sein kann aber nur, was gut gemacht ist. (...)

- **Gutes Design macht ein Produkt verständlich**
  Es verdeutlicht auf einleuchtende Weise die Struktur des Produkts. Mehr noch: Es kann das Produkt zum Sprechen bringen. Im besten Fall erklärt es sich dann selbst. (...)

- **Gutes Design ist unaufdringlich**
  Produkte, die einen Zweck erfüllen, haben Werkzeugcharakter. Sie sind weder dekorative Objekte noch Kunstwerke. Ihr Design sollte deshalb neutral sein, die Geräte zurücktreten lassen und dem Menschen Raum zur Selbstverwirklichung geben. (...)

- **Gutes Design ist ehrlich**
  Es lässt ein Produkt nicht innovativer, leistungsfähiger, wertvoller erscheinen, als es in Wirklichkeit ist. Es versucht nicht, den Verbraucher durch Versprechen zu manipulieren, die es dann nicht halten kann. (...)

- **Gutes Design ist langlebig**
  Es vermeidet, modisch zu sein, und wirkt deshalb nie antiquiert. Im deutlichen Gegensatz zu kurzlebigem Mode-Design überdauert es auch in der heutigen Wegwerfgesellschaft lange Jahre. (...)

- **Gutes Design ist konsequent bis ins letzte Detail**
  Nichts darf der Willkür oder dem Zufall überlassen werden. Gründlichkeit und Genauigkeit der Gestaltung sind letztlich Ausdruck des Respekts dem Verbraucher gegenüber. (...)

- **Gutes Design ist umweltfreundlich**
  Design leistet einen wichtigen Beitrag zur Erhaltung der Umwelt. Es bezieht die Schonung der Ressourcen ebenso wie die Minimierung von physischer und visueller Verschmutzung in die Produktgestaltung ein.

- **Gutes Design ist so wenig Design wie möglich**
  Weniger Design ist mehr, konzentriert es sich doch auf das Wesentliche, statt die

Produkte mit Überflüssigem zu befrachten.
Zurück zum Puren, zum Einfachen!“ [Vi13]

4.2 Prinzipien zur Gestaltung von Software

Die Internationale Normungsorganisation ISO (International Organization for Standardization) hat durch die Arbeitsgruppe 5 „Software-Ergonomie und Mensch-Computer-Dialoge“ des Technischen Komitees ISO/TC 159 „Ergonomie“, Unterkomitee SC 4 „Ergonomie der Mensch-System-Interaktion“, die Normenreihe DIN EN ISO 9241 „Ergonomie der Mensch-System-Interaktion“ (alter Titel: „Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten“) erarbeitet, um Hilfestellung für ein gutes Design von Software zu geben.

In der DIN EN ISO Norm 9241-11 „Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten Teil 11: Anforderungen an die Gebrauchstauglichkeit — Leitsätze (ISO 9241-11:1998)“ werden folgende drei Leitkriterien für die Gebrauchstauglichkeit von Software genannt [DIN99]: „

- **Effektivität**
  Die Genauigkeit und Vollständigkeit, mit der Benutzer ein bestimmtes Ziel erreichen. (...)
- **Effizienz**
  Der im Verhältnis zur Genauigkeit und Vollständigkeit eingesetzte Aufwand, mit dem Benutzer ein bestimmtes Ziel erreichen. (...)
- **Zufriedenstellung**
  Freiheit von Beeinträchtigungen und positive Einstellungen gegenüber der Nutzung des Produkts.“

Demnach wird auch in der DIN EN ISO Norm 9241-210 „Ergonomie der Mensch-System-Interaktion – Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme (ISO 9241-210:2010)“ die Gebrauchstauglichkeit/Usability von Software verstanden als „Ausmaß, in dem ein System, ein Produkt oder eine Dienstleistung durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um festgelegte Ziele effektiv, effizient und zufriedenstellend zu erreichen“ [DIN11]. Aufbauend auf diesen drei Leitkriterien werden in der DIN EN ISO Norm 9241-110 „Ergonomie der Mensch-System-Interaktion – Teil 110: Grundsätze der Dialoggestaltung (ISO 9241-110:2006)“ (welche DIN EN ISO Norm 9241-10 ersetzt) folgende Grundsätze für die Gestaltung und Bewertung von Benutzerschnittstellen zu einer Software aufgestellt [DIN06]: „

- **Aufgabenangemessenheit**
  Ein interaktives System ist aufgabenangemessen, wenn es den Benutzer unterstützt, seine Arbeitsaufgabe zu erledigen, d. h., wenn Funktionalität und Dialog auf den charakteristischen Eigenschaften der Arbeitsaufgabe basieren, anstatt auf der zur Aufgabenerledigung eingesetzten Technologie. (...)

- **Selbstbeschreibungsfähigkeit**
  Ein Dialog ist in dem Maße selbstbeschreibungsfähig, in dem für den Benutzer zu jeder Zeit offensichtlich ist, in welchem Dialog, an welcher Stelle im Dialog er sich befindet, welche Handlungen unternommen werden können und wie diese ausgeführt werden können. (...)

- **Erwartungskonformität**
  Ein Dialog ist erwartungskonform, wenn er den aus dem Nutzungskontext heraus vorhersehbaren Benutzerbelangen sowie allgemein anerkannten Konventionen entspricht. (...)

- **Lernförderlichkeit**
  Ein Dialog ist lernförderlich, wenn er den Benutzer beim Erlernen der Nutzung des interaktiven Systems unterstützt und anleitet. (...)

- **Steuerbarkeit**
  Ein Dialog ist steuerbar, wenn der Benutzer in der Lage ist, den Dialogablauf zu starten sowie seine Richtung und Geschwindigkeit zu beeinflussen, bis das Ziel erreicht ist. (...)

- **Fehlertoleranz**
  Ein Dialog ist fehlertolerant, wenn das beabsichtigte Arbeitsergebnis trotz erkennbar fehlerhafter Eingaben entweder mit keinem oder mit minimalem Korrekturaufwand seitens des Benutzers erreicht werden kann. (...)

- **Individualisierbarkeit**
  Ein Dialog ist individualisierbar, wenn Benutzer die Mensch-System-Interaktion und die Darstellung von Informationen ändern können, um diese an ihre individuellen Fähigkeiten und Bedürfnisse anzupassen.“

5 Mehr Nachhaltigkeit durch eine Verschiebung bzw. Auflösung der Grenze zwischen physischem Objekt und Software?

Grundsätzlich stellt sich die Frage, welches Entwurfsparadigma – das von physischen Objekten oder das von Software – bei Produkten wie der Apple Watch, wo sich die Grenze von physischem Objekt (Hardware) und Software nicht mehr bestimmen lässt, anzuwenden ist. Die ursprüngliche Leitidee bzw. Ausrichtung der Gestaltung von Software spiegelt die DIN EN ISO 9241-11 mit ihren drei Gebrauchstauglichkeit-Leitkriterien Effektivität, Effizienz und Zufriedenstellung wider. Diese Leitkriterien lassen sich auf die Erkenntnisse der Arbeitswissenschaften zurückführen und wurden vor dem Hintergrund von Software im Büro- und Produktionsbereich formuliert, d.h. die Frage einer Gebrauchstauglichkeit von Software war vor allem auf die Berufswelt von Nutzern hin orientiert [ReG13].

Analog zum Software Engineering wurde deshalb ein Usability Engineering (siehe beispielhaft Abbildung 2) etabliert, in dessen Mittelpunkt sogenannte Usability Tests ste-

hen. Usability Tests sollen Erkenntnisse darüber liefern, ob vorab festgelegte, quantitativ messbare Ziele in der Umsetzung bzw. im Betrieb einer Software auch erreicht wurden. Dies geschieht durch eine Beobachtung des Verhaltens der definierten Nutzergruppe(n) [Eic15]. Im Sinne einer (ansatzweisen) ständigen Verbesserung wird dieser Kreislauf solange durchlaufen, bis die messbaren Ziele erreicht werden.

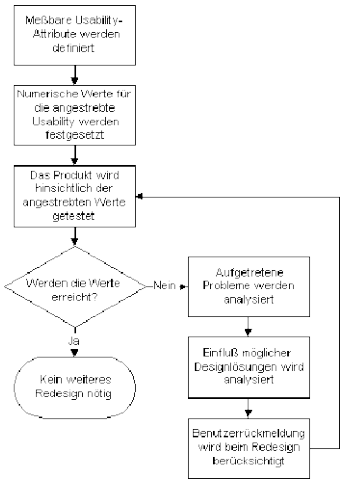


Abbildung 2: Usability Engineering (Quelle: [Eic15])

Nicht erst Produkte wie die Apple Watch zeigen aber, dass ein solch ausgestalteter Prozess eines Usability Engineering nicht adäquat die Gestaltung von Software (und auch Hardware) abbilden kann und deshalb der „als Produkteigenschaft definierten Qualitätsbegriff der *Usability* zu einem Qualitätsbegriff, welcher erst durch den Benutzungsprozess erfahrbar wird“ [Reg13], erweitert werden muss.

Diesem Umstand trägt die ISO bereits Rechnung, indem sie in diesem Zusammenhang von einem „Benutzererlebnis“ (englisch „User Experience“) spricht, also der „Wahrnehmungen und Reaktionen einer Person, die aus der tatsächlichen und/oder der erwarteten Benutzung eines Produkts, eines Systems oder einer Dienstleistung resultieren (...) User Experience umfasst sämtliche Emotionen, Vorstellungen, Vorlieben, Wahrnehmungen, physiologischen und psychologischen Reaktionen, Verhaltensweisen und Leistungen, die sich vor, während und nach der Nutzung ergeben“ [DIN11].

Damit macht die Verschiebungen der Grenze von physischem Objekt und Software deutlich, dass bei dem Entwurf von Produkten wie der Apple Watch in jedem Fall auf die Erfahrungen bzw. die Prinzipien des Gestaltens von physischen Objekten zurückgegriffen werden muss. Somit sollte auch ein ausgewogenes Verhältnis von kreativen und ingenieurwissenschaftlich abgeleiteten Erkenntnissen in den Entwurfsprozess Eingang finden. Ein rein ingenieurwissenschaftliches Vorgehen, wie es das Usability Engineering entsprechend dem Vorgehensmodell in Abbildung 2 vorschlägt, wird nicht zu den gewünschten Ergebnissen eines geforderten Benutzererlebnisses führen. Und das erfordert

wiederum einen *Perspektivwechsel* für Entwurfsprozesse in der Informatik bzw. gerade auch in der Umweltinformatik.

Es ist also eine *Konvergenz* der Prinzipien des Gestaltens von physischen Objekten (z.B. die 10 Thesen des Guten Designs nach Rams) und der Prinzipien des Gestaltens von Software (z.B. nach DIN EN ISO Normenreihe 9241) einzufordern, um so den gestalterischen Anforderungen im Sinne eines Benutzererlebnisses zu genügen. Und diese Konvergenz wiederum führt signifikant zu *mehr Nachhaltigkeit* von IKT und bzw. IKS oder auch ganzen IT-Ökosystemen, da das Stichwort „Nachhaltigkeit" einen festen Platz in dem Katalog der Anforderungen beim Entwurf von physischen Objekten hat, wie es z.B. auch die Thesen „Gutes Design ist langlebig" und „Gutes Design ist umweltfreundlich" insbesondere verdeutlichen [vgl. auch Arn14].

6 Ausblick

Die Frage, ob ein Perspektivwechsel in der Umweltinformatik durch ein mögliches Verschwinden der Grenze zwischen physischem Objekt und Software notwendig ist, kann eindeutig mit einem Ja beantwortet werden. Ein solcher Perspektivwechsel in der Umweltinformatik bietet zudem auch die Chance, die Nachhaltigkeit von IKT und IKS positiv beeinflussen zu können. Richtungsweisend dabei ist, dass die historische Trennung der Gestaltung von physischem Objekt und Software aufgehoben wird und eine Konvergenz der jeweiligen Gestaltungsprinzipien angestrebt wird. Dies ist auch ganz im Sinne der DIN EN ISO Normenreihe 9241 zur Ergonomie der Mensch-System-Interaktion zu sehen, da so das von der ISO geforderte Benutzererlebnis in einem nachhaltigen Sinne umgesetzt werden kann.

Literaturverzeichnis

[App04] Apple Inc.: iPod mini: Benutzerhandbuch, 2004 https://manuals.info.apple.com/MANUALS/0/MA734/de_DE/iPodminiUG.pdf, [2015-04-20].

[App09] Apple Inc.: iPhone Benutzerhandbuch: Für iPhone OS 3.1 Software, 2009 https://manuals.info.apple.com/MANUALS/0/MA616/de_DE/iPhone_iOS3.1_Benutzerhandbuch.pdf , [2015-04-20].

[App14] Apple Inc.: Apple Watch - Introducing Apple Watch, 30.09.2014, https://www.youtube.com/watch?v=y-waTi8BPdk [2015-04-20].

[App15a] Apple Inc.: Apple Watch ab 24. April in neun Ländern verfügbar, Apple Presseinformation 09. März 2015, http://www.apple.com/de/pr/library/2015/03/09Apple-Watch-Available-in-Nine-Countries-on-April-24.html [2015-04-20].

[App15b] Apple Inc.: Apple Watch Technologie, 2015 https://www.apple.com/de/watch/technology/ [2015-04-20].

[App15c] Apple Inc.: Designing for Apple Watch, 2015-03-09, https://developer.apple.com/library/prerelease/ios/documentation/UserExperience/Conceptual/WatchHumanInterface

Guidelines/index.html#//apple_ref/doc/uid/TP40014992-CH3-SW1 [2015-04-20].

[Arn13] Arndt, H.-K.: Umweltinformatik und Design - Eine relevante Fragestellung? In: Horbach, M. (Hrsg.): INFORMATIK 2013: Informatik angepasst an Mensch, Organisation und Umwelt (16.–20. September 2013, Koblenz, Germany), GI-Edition-Lecture Notes in Informatics (LNI), P-220, Gesellschaft für Informatik e.V., Bonn, 2013, S. 931-939.

[Arn14] Arndt, H.-K.: Big Data oder Grand Management Information Design? In: Plödereder, E./ Grunske, L./Schneider, E./Ull, D. (Hrsg.): INFORMATIK 2014: Big Data - Komplexität meistern (44. Jahrestagung der Gesellschaft für Informatik, 22.–26. September 2014, Stuttgart Germany), GI-Edition-Lecture Notes in Informatics (LNI), P-232, Gesellschaft für Informatik e.V., Bonn, 2014, S. 1947-1956.

[DIN99] Deutsches Institut für Normung e.V. (Hrsg.): DIN EN ISO Norm 9241-11 „Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten Teil 11: Anforderungen an die Gebrauchstauglichkeit — Leitsätze (ISO 9241-11:1998), Beuth Verlag, Berlin/Wien/Zürich, 1999.

[DIN06] DIN EN ISO Norm 9241-110 „Ergonomie der Mensch-System-Interaktion – Teil 110: Grundsätze der Dialoggestaltung (ISO 9241-110:2006), Beuth Verlag, Berlin/Wien/Zürich, 2006.

[DIN11] DIN EN ISO Norm 9241-210 „Ergonomie der Mensch-System-Interaktion – Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme (ISO 9241-210:2010), Beuth Verlag, Berlin/Wien/Zürich, 2011.

[Dud13a] Bibliographisches Institut GmbH (Hrsg.): Akuator, 2013, http://www.duden.de/rechtschreibung/Aktuator [2015-04-20].

[Dud13b] Bibliographisches Institut GmbH (Hrsg.): Krone, 2013, http://www.duden.de/rechtschreibung/Krone [2015-04-20].

[Eic15] Eichinger, A.: Usability Engineering, Universität Regensburg, http://www.uni-regensburg.de/Fakultaeten/phil_Fak_II/Psychologie/Doktoranden/absolventen/eichinger_armin/u-engineering.html [2015-06-15].

[Mos14] Mossberg, W.: It's Tim Cook's Apple Now, 09.09.2014, http://recode.net/2014/09/09/its-tim-cooks-apple-now/ [2015-04-20].

[OTW13] Obsolete Technology Website (Ed.): Old Macintosh - 1984 By Apple Computer, 2013 http://oldcomputers.net/macintosh.html [2015-04-20].

[PlP13] Platzner, M.; Plessl, C.: Verschiebungen an der Grenze zwischen Software und Hardware, Engineering Proprioception in Computing Systems Project (EPICS) 2013, http://www.epics-project.eu/publications/2014_plessl_automatismen.pdf [2015-04-20].

[ReG13] Reiterer, H.; Geyer, F.: C 6 Mensch-Computer-Interaktion. In: Kuhlen, R.; Semar, W.; Strauch, D. (Hrsg.): Grundlagen der praktischen Information und Dokumentation, 6. Aufl., Walter de Gruyter Verlag, Berlin, 2013, S. 413-440.

[Vit15] Vitsoe Ltd (Hrsg.): Dieter Rams: Zehn Thesen für gutes Design, 2015 https://www.vitsoe.com/de/ueber-vitsoe/gutes-design [2014-06-15]

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

From Ecological Modeling to Decentralized Optimization of Smart Power Grids

Michael Sonnenschein[1] Ute Vogel[2] Christian Hinrichs[3]

Abstract: In this contribution we want to sum up how computer science, ecological modeling and control of distributed power systems mutually influence. First we show, how ecological modeling has been influenced by the availability of faster computers and easier to use programming environments. Then we argue that adaption and emergent behavior in biological systems observed e.g. at the hand of microscopic models can be interpreted as an optimization process. This observation led to new optimization (meta-)heuristics in computer science and inspired multi-agent based optimization and control methods. We motivate why these methods are an interesting choice for finding new solutions to planning and control problems in future electrical power systems. As an example we sketch a completely distributed, agent-based planning and control system for a so-called dynamic virtual power plant. Finally we discuss some directions of future research in the domain of agent-based control of distributed systems.

Keywords: Ecological Modeling, Nature-inspired Heuristics, Multi-agent Systems, Electrical Power Systems

1 Introduction and Overview

For several years now, 'swarm intelligence' and 'nature-inspired heuristics' are considered to be a promising approach for optimizing large distributed systems. This is particularly the case for optimization problems in the electrical power system (see e.g. [Mc07, AEH09, Ka14]).

In this contribution we want to sum up how computer science, ecological modeling and control of distributed power systems mutually influence. First we give a short introduction into ecological modeling and show, how this discipline has been influenced by the availability of faster computers and easier to use programming environments: particularly these conditions allowed the creation of microscopic population models. Then we argue that adaption and emergent behavior in biological systems observed, e.g. at the hand of microscopic models, can be interpreted as an optimization process. This observation led to new optimization (meta-)heuristics in computer science.

Currently our electrical power supply system is redesigned from a few large coal-fired or nuclear power plants to a large set of relatively small distributed power plants based on re-

[1] University of Oldenburg, Department of Computing Science, Environmental Informatics, 26111 Oldenburg, Germany, sonnenschein@informatik.uni-oldenburg.de
[2] ute.vogel@uni-oldenburg.de
[3] christian.hinrichs@uni-oldenburg.de

newable energy. But power generation from photovoltaic and wind power plants is volatile and does not necessarily correlate spatially and temporally to the demand. This and other reasons cause a substantial increase of complexity in controlling these distributed units to guarantee grid stability and reliability of supply. After a very brief introduction into current challenges of our power system in section 4 we motivate why optimization heuristics and artificial intelligence methods are an interesting choice for finding new solutions to control problems in this context. In section 4.2 we show the enormous increase of research in this area in the last 10 years referring to Scopus queries. As an example in section 4.3 we sketch a completely distributed, agent-based control system for a so-called virtual power plant that was motivated by swarm behavior in nature. In the conclusion we discuss the advantages of such nature-inspired, distributed control methods but also their risks that lead to further research in this domain.

2 Ecological Modeling

Ecological systems are highly complex systems which are characterized by diverse and interacting (social) components. The objective of modeling, in particular of ecological modeling, consists in breaking this complexity down to a (formal) model which reduces the system to some application-specific core elements. The resulting abstract model can be analyzed and helps to gain better insights in the dynamics of the underlying system. With the development of Theoretical Biology, mathematical models were increasingly used: The resulting models are open for qualitative analysis, e.g. graphical representations help to uncover structural dependencies such as cause-effect-relations hierarchies and feedback cycles. Quantitative models, especially differential equations, allow the usage of models for estimating gaps in the collected environmental data or even to extrapolate the future development of the system [Lo25, Fo93]. With the availability of computers, software tools have been developed (Stella, et al.) to support the analysis and allow comfortable simulations. The abstraction level of such system dynamics models is very high: the diversity and heterogeneity of the original system is reduced to a small number of system state variables and parameters and the dynamics described by flows between the state variables. This very abstract modeling is justified in systems with a very high number of individual components as in such systems the aggregated behavioral deviations of individuals can be approximated by the average behavior using averaged parameters. Due to the high degree of abstraction, more detailed views to the system evolved: Individual-oriented models allowed for individual distribution models (IBDM) which split a population by an individual attribute, e.g. age, and model the resulting groups of individuals as separate state variables, but still represent abstracted views to an aggregated system's dynamics which cannot directly be observed in nature. This changed with increased computational power and availability of adequate software tools, e.g. [LS98]: Individual-oriented models (IBCM) go one step further by modeling the individuals of the system and their individual behavior itself, and hence, allow observing interactions between individuals as well as between individuals and environment [HH90, DRH94]. Simulated flocks of birds and fish schools are prominent examples that such models could be used for validating biological theories of emergence phenomena. Additionally, the concept of Cellular Automata [Ho88]

allows the micro-scale modeling of spatial processes and the integration of heterogeneous spatial conditions into models.

Common to all these Ecological Modeling approaches is that their objective is to gain insights in the behavior of biological and ecological systems. [HDP88]

3 Nature-Inspired Optimization

The micro-scale approach of individual-oriented modeling allows to model the individual behavior and to observe emergent effects on a macro-scale layer. In combination with the fact that the behavior of biological or ecological systems can be interpreted as an optimization to their environment, this observation has led to new optimization (meta-) heuristics, so-called nature-inspired optimization. Introductory textbooks on this subject are e.g. [Ya10, Br12]. They include several examples of applications as well as pseudo-code for algorithms.

Other than the ecological modeling, nature-inspired heuristics are applied to optimization problems in arbitrary disciplines. Applications range over a broad area of computational problems, e.g. network routing and network design, maximal independent set, minimum spanning tree, space partitioning [NBJ14]. So, there has been a change from applying computer technology for the understanding of natural systems to using patterns of natural behavior for solving computational expensive or intractable optimization tasks [Co96].

In the following, some examples of such nature-inspired optimization techniques are briefly introduced. We differentiate between heuristic approaches with central control and agent-based approaches which produce near-optimal solutions solely by interacting agents.

3.1 Heuristic Approaches

The structure and behavior of Artificial Neural networks (ANN) mimics biological neural networks. They consist of a set of interconnected (artificial) neurons which receive, process and transmit impulses. An artificial neuron is usually only capable of simple computational steps, i.e. the linear combination of inputs and the application of thresholds. The weights and thresholds which are used in the transfer and processing can be trained in order to map a set of inputs (patterns) to outputs [Mi12]. Hence they reflect a mere statistical relation in the input/output data and do not explain any structural aspects. On the global perspective, the realized input/output behavior of a trained neural network can be seen as an emergent process.

Evolutionary algorithms (EA) are one of the best-known nature-inspired optimization meta-heuristics: The evolutionary process which allows species or ecosystems to evolve gradually to an optimal adaption to its environment has inspired different variants of optimization techniques [FP95]. The basic idea which has been implemented in Genetic Algorithms (GA) is to view (non-optimal) solutions of an optimization task as individuals

of a population. Chromosomes code the description of the solution. The nature-inspired processes Crossover and Mutation produce candidates for the next generation of solutions by constructing a new combination of two parental chromosomes and by random changes in some attributes (genes) of a chromosome. The candidates which are the fittest related to an arbitrary objective function are chosen for the next generation. So the information about the solution space is distributed over the individuals and coded in their chromosomes, but the evaluation of this information takes place in the centralized selection process.

Particle Swarm Optimization (PSO) [UP11] has been inspired by swarms, herds, and flocks of animals. Similar to GA, a PSO starts with a random population of solutions. The individuals explore the solution space by moving around with individual velocity and direction. After each step, the individuals' solution qualities are evaluated and the position of the global best solution determined. Each individual adapts its direction, and hence its position in the next iteration, by taking its own best as well as the global best solution into account.

In this basic form, the behavior of the particle swarm cannot be viewed as an intelligent swarm [SLW12], as the propagation of the determined global best solution represents a central component. Variants of the PSO modify the propagation of the individual best solution to a local neighborhood and hence can be viewed as intelligent swarm. Furthermore, the knowledge propagation about the known best solution changes to a decentralized process, which opens up the possibility of a completely distributed implementation.

3.2 Agent-Based Methods

In the micro-scale approach of ecological modeling it often can be observed, that the individuals' behavior leads to a global optimum of the overall ecological or biological system. The behavior of fish schools [Re87] illustrates this idea very well: The movement of a fish school completely depends on the simple behavior of the single fishes, their restricted perception of their environment, and their interaction. Each fish's behavior can be reduced to three rules: 1. Avoid collisions 2. Move with the same velocity and direction as other fishes, and 3. Head for fishes which might be joined in the school. Simulations show that these rules can explain the school's movement, which represents an optimal protection strategy. Such systems are called *self-organizing*, as there is no central entity that controls the system's progress, and the global system behavior purely emerges from local interactions. This idea of independent agents with simple behavioral rules and only local knowledge about their environment in combination with other species behavioral patterns has been adopted by agent-based heuristic optimization methods: The behavior of biological ants and bees inspired heuristics for determining short routes by the Ant Colony Optimization (ACO) and for optimizing routing protocols by the Bee Colony Algorithm, resp. [GP10]. In these heuristics the individuals can be viewed as agents in a multi-agent system. Each individual has its own attributes and perception of its local environment and is interacting and communicating with other individuals in order to fulfill its own objectives. Such a system fulfills the definition of swarm intelligence as 'ability to act in a coordinated way without the presence of a coordinator or of an external controller' [SLW12].

4 Challenges in Electrical Power Systems

The purpose of an electrical power system is clearly to provide us with electrical power in the most reliable, economic, environmental friendly, and socially acceptable manner. Obviously these four goals can't be maximized simultaneously, so political and societal weights for these four goals are necessary to design the best possible electrical power system based on the available technology. These weights can be implemented by legal frameworks, market regulations or financial incentives.

Currently, in Germany and in many other countries we undergo a system change originated by weighting up environmental aspects. In combination with emerging new technologies (wind turbines, photovoltaic systems, battery storage) this leads to a major redesign of our power supply system replacing large, nuclear or coal-fired power plants by small renewable energy sources (see e.g. [DII12] for a required scenario of the German power system in the year 2050). In consequence not only the traditional role models and business cases of utility companies are changing, e.g. by private ownership of roof-top solar plants or small co-generation plants, but also accounting systems and control systems of this more and more distributed power supply system are being reconsidered significantly. Additionally the power grid has to be redesigned to adapt to the modified power flow in the grid caused by relocating power plants (see [5014] for the planned development in Germany which is the subject of controversial political discussion).

4.1 Operation and Control Aspects in Power Systems

Many comprehensive introductions into the operation of power systems are available (see e.g. [Me06, Sc12]). Of course in this contribution we have to view things from on a very high level view. In simple terms, apart from the need of matching of supply and demand operation and control of the power system has at least four technical goals to ensure stability:

- The frequency has to be 50Hz at any time (for the continental European power grid) with a maximal deviation of 0.2Hz.
- The voltage band on each voltage level must not be violated (e.g. 210V. . . 250V for the low voltage grid) at any point of the power grid.
- The operating limitations of all active and passive components (e.g. lines, transformers) of the power systems have to be respected.
- The quality of frequency and voltage, e.g. the level of short-term fluctuations and the harmonic-content, has to be optimized.

These are hard constraints to avoid power outages or damage of components. Additionally, from a macro-economic perspective, operation and control of the power system should minimize overall operational cost and also minimize emissions or other type of waste from the operation of nuclear or fossil-fuel power plant. In practice, cost-optimization of power

generation refers to single owners or owner communities and is limited by the market conditions. Environmental aspects are integrated into the operation by market rules, e.g. the EEG in Germany [Bu14], that guaranty priority of renewable energy.

So, operation and control of the power supply system includes a highly complex, multi-criteria optimization problem, integrating a huge amount of units (e.g. power plants or controllable loads) with many technical constraints and hard real-time aspects. This complex problem is solved by decomposition into several subproblems. Subproblems can be defined on the one hand by subtasks, e.g. frequency control or local voltage control, and on the other hand by identifying local control circuits, e.g. related to the grid topology, or economical balancing groups.

But with the redesign of our power supply system these operation and control problems becomes even more complex, requiring a 'smart grid'. For example, power generation from photovoltaic and wind power plants is volatile and does not necessarily correlate spatially and temporally to the demand. Moreover, prognosis of power generation from renewable energy comes with some remaining uncertainty. From a technical perspective the system's size also becomes problematic, as the number of power plants to be controlled increases by some orders of magnitude. Furthermore, power plants are now connected to the power grids at all voltage levels – not only at the high voltage and highest voltage level as in the past. To handle these challenges, first approaches aim at developing new decomposition methods – or vice versa aggregation methods for small power plants: virtual power plants (VPPs) [AP97, PRS07] and microgrids [Ha14]. Both attempts combine the operation of small power-plants whereby VPPs mainly act at the market whereas microgrids aim at a local supply-demand matching. But the increasing complexity also motivates research in new optimization and control methods, which will be discussed in the following sections.

4.2 Nature-Inspired Methods

Due to the substantially increased complexity of the power system, optimization heuristics and artificial intelligence methods are an interesting choice for finding new solutions. Thus, research in nature-inspired approaches for power system operation and control gained significant momentum in the past decades. Figure 1 shows the number of publications in this area for the last thirty years, based on a search in the Scopus bibliographic database.[4] Within these publications, most of the research focuses on artificial neural networks as well as evolutionary algorithms, beginning in the late 1980's. Ecologically inspired approaches such as ant/bee/bacterial foraging strategies or firefly/cuckoo search algorithms started appearing in the power systems in the late 1990's (summarized as "other" in the figure). A remarkable milestone was the adoption of particle swarm optimization in the power systems community in the early 2000's, leading to a rapid increase of research in this context.

While these approaches all focus on heuristic optimization, the field of multi-agent systems is more related to coordination and planning problems. Starting already in the 1990's with

[4] `http://www.scopus.com`, accessed June 20, 2015

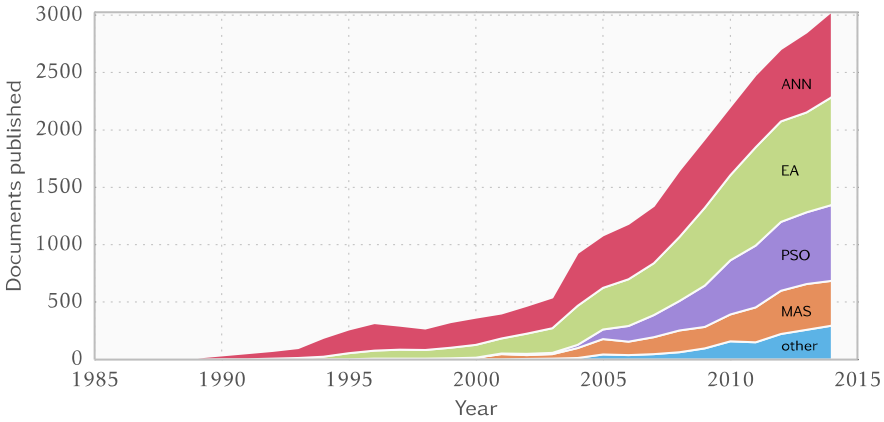


Fig. 1: Scopus query results for research in nature-inspired approaches with application to power systems in the last 30 years, displayed as stacked chart grouped by topic (ANN = artificial neural networks, EA = evolutionary algorithms, PSO = particle swarm optimization, MAS = multi-agent systems, other = ant/bee/bacterial foraging, firefly/cuckoo search algorithms and similar approaches).

only few publications per year in the power systems community, a significant increase of research volume can be seen nowadays in this area, see Fig. 2. This is supported by

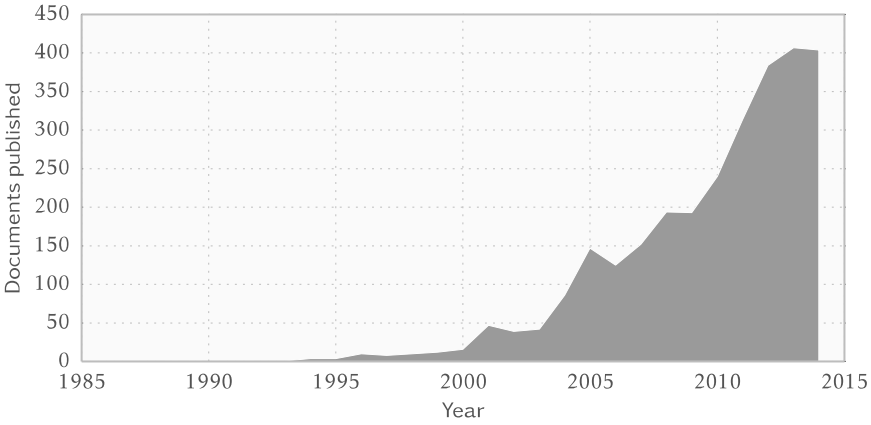


Fig. 2: Scopus query results for research in decentralized nature-inspired approaches with application to power systems in the last 30 years.

the trend towards complex intelligent distributed systems in power systems operation and control, as already outlined in Sect. 4.1. For instance, a research agenda from the point of view of computer science has recently been published in [Ra12], which particularly motivates distributed and agent-based solutions.

4.3 Agent-Based Approaches in Power Systems

As a practical example for an agent-based approach in power systems operation and control, let us consider the balancing of supply and demand. This is necessary for economical reasons and with respect to the first technical goal mentioned in Sect. 4.1. Frequency deviations in the power grid are caused by an imbalance between consumed and produced power. If more power is produced than consumed, the frequency increases and under reverse conditions it decreases. Simplified, this currently results in three tasks of the power supply system:

- Balance between demand and supply has to be achieved first by direct contracts between consumers and producers or at the power market. The latter is enabled by a complex set of products traded in this market [Eu14].
- The operation of controllable power plants (as well as controllable loads such as heat-pumps) has to be planned within groups of plants having jointly sold power products or load profiles in their supply area.
- During execution of the operational plan a system of power reserves has to be available to react instantaneously to unforeseen deviations from forecasts and other events as outage of a power plant. This task is primarily assigned to the transmission grid operator [Eu09].

Let us focus on a the second aspect: the planning of the operation of controllable power plants within a group responsible for delivering a specific power product or load-profile. This planning problem is very well known as the unit-commitment-problem. Established solutions exist for small groups of large controllable plants; for an overview see e.g. [Pa04]. For several years also nature-inspired optimization heuristics are considered to tackle this problem – see e.g. [KBP96] for an early approach. But with the redesign of the power system new challenges arise. For instance, the number of plants to be controlled increases, so highly scalable solutions are required. Further, the members of virtual power plants may change quite often, which calls for easily adaptable optimization methods. Because units in such a virtual power plant can have different owners, each unit also follows its own optimization goals, yielding hybrid multi-objective optimization problems. Finally, owners of power plants might have privacy restrictions regarding information on the operating status of their units. So, optimization might be restricted to limited, abstract state information of units.

Thus, a dynamic agent-based approach seems appropriate for this task. As an example, a suitable approach exists with the concept of Dynamic Virtual Power Plants (DVPP), which is summarized briefly as follows (cf. [Ni12, So15]). In a self-organized way coalitions of plants form with respect to concrete products in an energy market. After delivering a product, a coalition dissolves and the former participating units can then self-determinedly join the formation process of other coalitions for subsequent tradeable energy products. In particular, this comprises the following subprocesses:

1. *DVPP setup:* Plants are aggregated to DVPPs by coalition formation, such that the members of each DVPP agree upon trading a specific power product in the market (e.g. a certain block product in an electricity spot market). Bids for these products are then placed in the market by the respective DVPPs.
2. *Internal scheduling:* After a successful bid, a DVPP is obliged to deliver the power product. For this, the members of the DVPP have to be scheduled within their individually defined degrees of freedom. This is done prior to the actual delivery of the product in a predictive scheduling process which uses mathematical abstractions of the devices' feasible scopes of action.
3. *Continuous scheduling:* To compensate for unforeseen changes or forecast errors, a continuous scheduling is performed during the delivery of the product. Here, the units' schedules are adapted such that product delivery is not endangered.
4. *Payoff division:* Subsequently, the revenues gained from product delivery are distributed among the DVPP members, taking the actual commitment of the units during delivery into account.

In addition to the dynamic nature of this concept, all optimization and control tasks outlined above are realized with fully distributed approaches. Each participating plant is represented by an agent in the system, and the autonomy of the participants is preserved by employing self-organization strategies in the subprocesses. Some results from a preliminary implementation of this concept can be found in [As14].

5 Lessons Learned

In this article, we have sketched how computer science is related to the domain of ecological modeling and to the domain of power system's operation and control. Considered historically, computer science enabled new and much more detailed methods for modeling ecological systems, e.g. by individual-oriented models. This resulted in new insights in the behavior of natural systems. Vice versa, a deeper knowledge of self-organization and adaption in ecological systems inspired new optimization methods as e.g. evolutionary algorithms or methods based on the self-organization of social insects. Although developed somewhat in parallel, individual-oriented modeling in ecology has clearly a strong relationship to multi-agent systems, which are an excellent framework for distributed optimization and distributed control.

As technical systems become more and more complex and distributed, components in these systems become more 'intelligent' by advanced micro-controllers. Besides, communication channels between distributed components become self-evident. So distributed agent-based methods are promising candidates for optimization and control aspects in such technical systems. Electrical power systems are a prime example for this. By replacing a few large fossil-fired or nuclear power supply units by many small wind turbines and photovoltaic systems widely distributed around the power grid, electrical power systems become an even more complex and distributed socio-technical systems.

Self-organized agent-based systems promise in particular high adaptability, fault-tolerance regarding single components, scalability, and – not to forget – privacy of state information of components. But let us reconsider the difficulties that arise with this possible paradigm shift – particularly with regard to the criticality of the infrastructure that has to be controlled. From the viewpoint of ecological and individual-oriented modeling, self-organizing systems usually have been constructed by induction: By analyzing existing systems in nature, interactions among the systems' individuals could be identified and subsequently be adopted to optimization problems. This resulted in widely applicable concrete approaches, such as meta-heuristics.

Regarding the current and upcoming challenges in the optimization and control of complex critical infrastructures however, these generic solutions are often not sufficient any more. Instead, specialized methods and architectures with guaranteed properties are required, where a deductive approach could be more suitable. In such an approach, a system would be constructed with a concrete goal in mind, but based on an abstract yet rigorous methodology to model the system elements and their interactions towards the desired emerging behavior. For instance, in the context of communication networks, an example of such a methodology has been proposed in [PB05]. We believe that this kind of approach will become more important in the future. Although, due to their complex internal dynamics, distributed algorithms and agent-based methods are more demanding with respect to proofs and guarantees on their intended behaviour. Also, agent-systems are in risk to be vulnerable by malicious agents. For an actual adoption in critical infrastructures, these are significant aspects.

In summary, new research challenges arise for design methods for distributed optimization and control methods, especially with a focus on provable guaranteed behavior. In this sense, also the integration of security aspects into agent-based frameworks play a crucial role, see [Ne14]. Finally, high-level standards for complex communication issues in multi-agent systems controlling power system's components will be subject to intensive work.

References

[5014] 50Hertz Transmission GmbH; Amprion GmbH; TenneT TSO GmbH; TransnetBW GmbH: , Netzenwicklungsplan Strom, 2014.

[AEH09] AlRashidi, M. R.; El-Hawry, M. E.: A Survey of Particle Swarm Optimization Applications in Electric Power Systems. IEEE Transactions on Evolutionary Computing, 13(4):913, 2009.

[AP97] Awerbuch, Shimon; Preston, Alistair M., eds. The Virtual Utility: Accounting, Technology & Competitive Aspects of the Emerging Industry, volume 26 of Topics in Regulatory Economics and Policy. Kluwer Academic Publishers, 1997.

[As14] Astrid Nieße; Sebastian Beer; Jörg Bremer; Christian Hinrichs; Ontje Lünsdorf; Michael Sonnenschein: Conjoint Dynamic Aggregation and Scheduling Methods for Dynamic Virtual Power Plants. In (Ganzha, Maria; Maciaszek, Leszek A.; Paprzycki, Marcin, eds): Proceedings of the 2014 Federated Conference on Computer Science and Information Systems. volume 2 of Annals of Computer Science and Information Systems. IEEE, pp. 1505–1514, 2014.

[Br12] Brownlee, Jason: Clever Algorithms: Nature-Inspired Programming Recipes. lulu.com, 2012.

[Bu14] Bundesministerium für Wirtschaft und Energie: Gesetz zur grundlegenden Reform des Erneuerbare–Energie–Gesetzes und zur Änderung weiterer Bestimmungen des Energiewirtschaftsrechts. BGBl. Teil I Nr. 33/2014, 2014.

[Co96] Colorni, A.; Dorigo, M.; Maffioli, F.; Maniezzo, V.; Righini, G.; Trubian, M.: Heuristics from Nature for Hard Combinatorial Optimization Problems. International Transactions in Operational Research, 3(1):1–21, 1996.

[DII12] DLR; IWES; IFNE: Langfristszenarien und Strategien für den Ausbau der erneuerbaren Energien in Deutschland bei Berücksichtigung der Entwicklung in Europa und global. Bundesministerium für Umwelt, Naturschutz, Bau und Reaktorsicherheit, 2012.

[DRH94] DeAngelis, D. L.; Rose, K. A.; Huston, M. A.: INDIVIDUAL-ORIENTED APPROACHES TO MODELING ECOLOGICAL POPULATIONS AND COMMUNITIES. In (Levin, Simon A., ed.): Frontiers in Mathematical Biology, volume 100 of Lecture Notes in Biomathematics, pp. 390–410. Springer, 1994.

[Eu09] European Network of Transmission System Operators for Electricity: , Policy 1: Load-Frequency Control and Performance, 2009.

[Eu14] European Power Exchange: , EPEX Sport Operational Rules, 2014.

[Fo93] Forrester, Jay W.: System Dynamics and the Lessons of 35 Years. In (Greene, Kenyon B. de, ed.): A Systems-Based Approach to Policymaking, pp. 199–240. Springer US, Boston and MA, 1993.

[FP95] Fonseca, Carlos M.; Peter, J.: An Overview of Evolutionary Algorithms in Multiobjective Optimization. In (Beyer, H.-J., ed.): Evolutionary Computation, volume 3:1, pp. 1–16. MIT Press Cambridge, MA, USA, 1995.

[GP10] Gendreau, Michel; Potvin, Jean-Yves: Handbook of Metaheuristics, volume 146. Springer US, Boston and MA, 2010.

[Ha14] Hatziargyriou, Nikos, ed. Microgrids: Architecture and Control. Wiley - IEEE, 2014.

[HDP88] Huston, Michael; DeAngelis, Donald; Post, Wilfred: New Computer Models Unify Ecological Theory. BioScience, 38(10):682–691, 1988.

[HH90] Hogeweg, P.; Hesper, B.: Individual-oriented modelling in ecology. Mathematical and Computer Modelling, 13(6):83–90, 1990.

[Ho88] Hogeweg, P.: Cellular automata as a paradigm for ecological modeling. Applied Mathematics and Computation, 27(1):81–100, 1988.

[Ka14] Karaboga, D.; Gorkemli, B.; Ozturk, C.; Karaboga, N.: A comprehensive survey: artificial bee colony (ABC) algorithm and applications. Artif Intell Rev, 42:21–57, 2014.

[KBP96] Kazarlis, S. A.; Bakirtzis, A. G.; Petridis, V.: A genetic algorithm solution to the unit commitment problem. IEEE Trans. on Power Systems, 11(1):83–92, 1996.

[Lo25] Lotka, Alfred J.: Elements of Physical Biology. Williams and Wilkins Company, 1925.

[LS98] Lorek, Helmut; Sonnenschein, Michael: Object-oriented support for modelling and simulation of individual-oriented ecological models. Ecological Modelling, 108(1-3):77–96, 1998.

[Mc07] McArthur, S.D.J..; Davidson, E. M.; Catterson, V. M.; Dimeas, A. L.: Multi-Agent Systems for Power Engineering Applications – Part I: Concepts, Approaches, and Technical Challenges. IEEE Transactions on Power Systems, 22(4):1743–1752, 2007.

[Me06] Meier, Alexandra von: Electric Power Systems: A Conceptual Introduction. Wiley-IEEE Press, 2006.

[Mi12] Mikhailov, Alexander S.: From statistics to neural networks: Theory and pattern recognition. Springer, [S.l.], 2012.

[NBJ14] Navlakha, Saket; Bar-Joseph, Ziv: Distributed information processing in biological and computational systems. Communications of the ACM, 58(1):94–102, 2014.

[Ne14] Neureiter, C.; Eibl, G.; Engel, D.; Schelgel, S.; Uslar, M.: A concept for engineering smart grid security requirements based on SGAM models. Computer Science – Research and Development, 2014.

[Ni12] Nieße, Astrid; Lehnhoff, Sebastian; Tröschel, Martin; Uslar, Mathias; Wissing, Carsten; Appelrath, Hans-Jürgen; Sonnenschein, Michael: Market-based self-organized provision of active power and ancillary services: An agent-based approach for Smart Distribution Grids. In: Complexity in Engineering (COMPENG). 2012.

[Pa04] Padhy, N. P.: Unit commitment - a bibliographical survey. IEEE Trans. on Power Systems, 19(2):1196–1205, 2004.

[PB05] Prehofer, Christian; Bettstetter, Christian: Self-Organization in Communication Networks: Principles and Design Paradigms. IEEE Communications Magazine, 43(7):78–85, 2005.

[PRS07] Pudjianto, D.; Ramsay, C.; Strbac, G.: Virtual power plant and system integration of distributed energy resources. Renewable Power Generation, 1(1):10–16, 2007.

[Ra12] Ramchurn, Sarvapali D.; Vytelingum, Perukrishnen; Rogers, Alex; Jennings, Nicholas R.: Putting the 'Smarts' into the Smart Grid: A Grand Challenge for Artificial Intelligence. Commun. ACM, 55(4):86–97, 2012.

[Re87] Reynolds, Craig W.: Flocks, herds and schools: A distributed behavioral model. In (Stone, Maureen C., ed.): Proceedings of SIGGRAPH '87. volume Vol. 21, No. 4 of Computer Graphics Ser, Association for Computing Machinery, New York, pp. 25–34, Oct. 1987.

[Sc12] Schwab, Adolf J.: Elektroenergiesysteme: Erzeugung, Transport und Verteilung elektrischer Energie. Springer, 2012.

[SLW12] Sun, Jun; Lai, Choi-Hong; Wu, Xiao-Jun: Particle swarm optimisation: Classical and quantum perspectives. Chapman and Hall/CRC numerical analysis and scientific computing. CRC Press, Boca Raton and FL, 2012.

[So15] Sonnenschein, Michael; Hinrichs, Christian; Nieße, Astrid; Vogel, Ute: Supporting Renewable Power Supply Through Distributed Coordination of Energy Resources. In (Hilty, Lorenz M.; Aebischer, Bernard, eds): ICT Innovations for Sustainability, volume 310 of Advances in Intelligent Systems and Computing, pp. 387–404. Springer International Publishing, 2015.

[UP11] Urade, Hemlata S.; Patel, Rahila: Article: Study and Analysis of Particle Swarm Optimization: A Review. IJCA Proceedings on 2nd National Conference on Information and Communication Technology, NCICT(4):1–5, 2011.

[Ya10] Yang, Xin-She: Nature-Inspired Metaheuristic Algorithms. Luniver Press, 2 edition, 2010.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Environmental and Energy Informatics: Polish Approaches in Eco-innovation

Pawel Bartoszczuk[1]

Abstract: The goal of this paper is to present eco-innovation implementation in Poland and in some European countries. I assess factor influencing on ecoinnovation in small medium enterprises. Both current and expected energy and material prices are particularly important for pushing firms to reduce pollution, decrease hazardous substances and increase recyclability of products. Cost savings are an important main incentive for reducing energy and material use, and for eco-innovation. Customer requirements are very important source for eco-innovations, particularly with regard to products with improved environmental performance and process innovations that increase material efficiency, reduce energy consumption and waste.

Keywords: ecoinnovation, energy, environmental regulation, Poland

1 Introduction

Definition of eco-innovation was demonstrated by Kemp and Pearson (2008). An environmental innovation, according to them, has been defined as "a new or significantly improved product (good or service), process, organizational method or marketing method that creates environmental benefits compared to alternatives. The environmental benefits can be the primary objective of the innovation or the result of other innovation objectives. The environmental benefits of an innovation can occur during the production of a good or service, or during the after sales use of a good or service by the end user" (Kemp and Pearson, 2008). In what follows is a catalog of environmental benefits that an eco-innovation could have produced either with the firm or from the after sales use of a product by the user for which surveyed companies should say whether this benefit has occurred or not. Concerning environmental product innovations, ZEW (Zentrum für Europäische Wirtschaftsforschung – Centre for European Economic Research) econometric results show that present regulations are only effective for reductions of air, water, soil and noise emissions but not suitable for energy consumption and recycling (Horbach, 2009). The companies expect a growing importance of future regulations for all product innovations. In all considered environmental innovation areas future regulations already trigger eco- innovations.

Environmental regulation and individual environmental policy instruments (specially, soft regulation) are significant drivers particularly for eco-product innovations (Cleff and Rennings, 1999, Horbach, 2008 and Rehfeld et al., 2007). Other researchers like Green

[1] Warsaw School of Economics, Enterprise Institute, pbarto1@sgh.waw.pl

et al. (1994) imply that firms implement environmental product innovation to obey with existing and anticipated legal requirements. Horbach (2008) finds a significant positive influence of subsidies on environmental product ecoinnovation. Kammerer (2009) demonstrate that a high level of regulatory stringency incentivises companies to implement environmental product innovations which are quite novel to the firm, but this result cannot be corroborated when these innovations are new to the market. Finally, Horbach et al. (2012) confirm a high importance of expected future regulations for all environmental product innovations.

Defining green innovation is not an easy task although several attempts have been made in the literature (Carrillo-Hermosilla et al., 2010). Klemmer et al. (1999) determined the environmental innovations as a subset of innovations that lead to an improvement of ecological equality. Chen et al. (2006, p.332) defined green innovation technologies that are involved in energy-saving, pollution-prevention, waste recycling, green product designs, or According to Halila and Rundquist (2011), the term, eco-innovation (or environmental innovation, green innovation or sustainable innovation), is often used to identify those innovations that contribute to a sustainable environment through ecological improvements. Eco-innovations are defined by Beise and Rennings (2003) as applications consist of new or modified processes, techniques, practices, systems and products to avoid or reduce environmental harms.

The current paper consist of 4 chapters, including introduction. In the second chapter I present results of ecoinnovation developments in Poland, in the third I present the topic of energy in Poland and the forth section presents the conclusions.

2 Ecoinnovation development in Poland

Poland, since its accession to the EU in 2004, enabled to utilize EU funds, reduce the backlogs in transportation infrastructure and environmental protection and at the same time builds a strong and stable economy (Kassenberg et. al, 2011). The state of the natural environment has significantly improved, while the resources productivity and energy intensity unfortunately have increased. Despite that, Polish productivity indexes per Gross Domestic Product still remain below expectations- less than average of the EU countries (Kassenberg et. al, 2011). During the last 20 years, energy consumption remained stable in spite of significant Gross Domestic Product (GDP) growth, due to energy efficiency improvement and changing the structure of economy. Nevertheless, the energy intensity index is still 2-3 times lower than the EU-27 average. In Poland, almost half (47%) of the companies surveyed by **Eurobarometer** stated that material costs represented 50% or more of their total production value. or more of their total production value (Flash EB No 315,2011). Companies in Poland were the most likely to have introduced a new or significantly improved eco-innovative product or service, production process or organisational method in the past two years (63%); companies in Hungary were the least likely to have done so (27%). Regardless of the fact that the country policy in the area of eco-innovation misses synergy, the eco-innovations have been ad-

dressed via national policy strategies on environmental protection, product policy, energy efficiency in buildings, etc. The interest of Small and Medium Enterprises (SME) in eco-innovation is slowly growing, especially in relation to cost reduction possibilities, due to notably reduction of energy consumption and decreasing expenditures related to pollutant emissions (Kassenberg et. al, 2011).

The country also has outstanding examples of eco-innovations in energy and water management, hazardous waste treatment, solar energy, green banking and coke industry and a number of eco-innovation related programs and initiatives within various clusters. On the other hand, eco-innovation does not constitute a driving force for new business opportunities in Poland. Development of eco-innovations in Poland is significantly hindered by numerous barriers. The most important is underestimation of the innovative potential as a growth driver. This innovative potential is understood here as an ability to create new technological and organizational solutions, increase labor productivity and improve resource efficiency with relation to work, capital, energy and materials. Eco-innovations in particular, remain outside the interest of decision makers policies which determine Poland's expectations towards the EU and future Structural Fund allocations for development objectives. As a consequence, Polish and some other countries' policies focus on clearing infrastructural backlog and accelerating extensive economic growth, ignoring future changes in the significance of the respective sectors and the barriers that limit the current growth model, which is based strongly on price, rather than quality, competition.

Another barrier comprises the big focus on EU funds availability and their fast utilization. The approach that money should be utilized fast, considerably limits the debate on how to optimize EU funds absorption. Poland needs stable mechanisms (similar to Finnish solutions) that will instigate the development of new, competitive structures of the Polish economy. There should be more focus on the quality of implemented ideas, their better selection and support for truly innovative economic initiatives.

The EU funds provide new possibilities for financing various initiatives at universities and colleges. Unfortunately, due to excessive formalization, more often the true aim of these activities is lost, while bureaucracy is increasing. A lot of attention is provided to the accounting side of projects, although it should be focused on how to fulfill the planned objectives and ensure durable benefits.

Lack of well-qualified and skillful specialist constitutes another large barrier to development of eco-innovations in Poland. This is result of, first and foremost, from a poor and ineffective system of education. Graduates of technological studies are in minority and staff available to work on eco-innovations is limited. There is a large difference between the business sphere, which is very innovative but still on a basic, everyday level, and administration and science. Due to insufficient information, access to eco-innovative solutions developed by the academic and science sector is limited. Providing adequate support for innovations, or eco-innovations in particular, constitutes a challenge. Poland has a serious problem with innovativeness as such, in its broad meaning.

Positive, albeit slow, changes are observed in Poland in the area of eco-innovations.

They do not immediately lead to a significant increase in the value of the eco-innovativeness scoreboard index. The Eco-Innovation Scoreboard (Eco-IS) is the first tool to assess and illustrate eco-innovation performance across the 27 EU Member States. The scoreboard aims at capturing the different aspects of eco-innovation by applying 16 indicators grouped into five thematic areas: eco-innovation inputs, eco-innovation activities, eco-innovation outputs, environmental outcomes and socio-economic outcomes. It thereby shows how well individual Member States perform in different dimensions of eco-innovation compared to the EU average and presents their strengths and weaknesses. The Eco-IS complements other measurement approaches of innovativeness of EU countries and aims to promote a holistic view on economic, environmental and social performance. Compared with 2011 Poland rank moved from 27th to the 25th position. The increase occurred in the area of eco-innovation output. (Eco-innovation in Poland: 2012 update)

There exist several examples of eco-innovation measures in Poland:

- Successful continuation of the program GreenEvo (Green Technologies Accelerator). This initiative of the Ministry of Environment supports Polish eco-innovators, mostly small and medium enterprises (SMEs) transferring technologies all around the world. In 2011 alone, Green Evo participants revenues increased by 31% on average, and their export revenues soared by 58%. What is more 86% of companies made a trade offer to foreign customers, and 50% of them have signed distribution agreements with foreign partners (Kassenberg et. al 2011).
- Establishing a number of eco-innovation oriented clusters e.g. Silesian Cluster of Environmental Technologies, Baltic Eco-energy cluster and Clean Energy Cluster of Małopolska and Podkarpackie (Eco-inovation in Poland: 2012 update.
- Participation of Poland in the Environmental Technologies Verification (ETV) Pilot Program of the European Union – a scheme supporting market uptake of eco-innovative technologies (Kassenberg et. al, 2011).

Despite these positive trends, development of eco-innovations in Poland is still hindered by a number of barriers. Eco-innovations still are giving the impression as "end of pipe, environmental protection technologies" rather than a cross- cutting innovations (Eco-innovation observatory 2012). Transition to a low-carbon economy is perceived as a threat in companies and SMEs. Implementation of stringent environmental regulations is seen solely as a cost and not as an opportunity for building an innovative and competitive economy. Awareness on the benefits resulting from implementation of eco-innovative technologies among entrepreneurs and consumers in general is relatively low. Entrepreneurs tend to invest in cheapest technologies allowing them to achieve the commercial goal or meet the minimum legislative requirement. Many entrepreneurs and research organization fail to see benefits from cooperation.

Lack of well-qualified and experienced labor force constitutes another large barrier to development of eco-innovations in Poland. This arises from, first and foremost, from a not effective system of education. Excessive number of humanists, lawyers or sociolo-

gist graduate from Polish universities each year, but there are too little engineers, and natural scientists, mathematicians etc. Staff available to work on eco-innovations is limited. There is a large difference between the business sphere, which is very innovative but still on a basic, everyday level, and administration and science. Due to insufficient information, access to eco-innovative solutions developed by the academic and science sector is limited. Providing adequate support for innovations, or eco-innovations in particular, constitutes a challenge. Poland has a serious problem with overall innovativeness as such, in its broad meaning.

The main barrier for Polish companies is the lack of sufficient capital to invent and implement eco-innovation. Additionally, financial institutions face a significant risk connected with involvement in eco-innovative projects. This risk is related more too technological issues, i.e. the possibility to achieve expected parameters, than to financial ones. There are no sufficient funds to verify proposed solutions, either in a semi-technical form or as a pilot solution. This relates especially to a project finance situation, where a company is established only to implement a particular investment. If a project has undergone preliminary verification, banks or other financial institutions are more likely to provide a loan, as such verification reduces their risk. Venture capital funds are also lacking. The establishment of the National Centre for Research and Development and development in 2007 and clusters should improve that situation.

3 Production of energy in Poland

Production of primary energy in Poland is mainly based on fossil fuels. The main prime energy source is still hard coal and lignite, which cover 56% of the demand. Crude oil also has a significant share of 25%. Even to maintain current levels of energy generation, Poland needs to invest huge amounts into energy generation capacity (between EUR 41 billion and EUR 98,5 billion by 2020 and factually upgrade or redesign its entire energy system:

- almost 85% of electricity is produced from coal
- two thirds of the installed coal capacity is older than 30 years
- almost 20% (7 gigawatt,) of the current generation capacity have to be phased out by 2015

However, we can observe a positive trend in the growth of renewable energy sector in Poland. The most important source is wind energy. According to Energy Regulatory Office (URE), in 2012, there existed 663 wind plants in Poland of a total capacity of 2341 MW. Most wind farms are located in the north-western Poland, not far from Baltics Sea. The leader is the Zachodniopomorskie Province (closest to the German border-716.8 MW), followed by the Pomorskie Province (246.9 MW) and the Wielkopolskie Province (245.3 MW). The current share of wind energy in total renewable electricity of origin is 57.6%. It ranked first among renewable energy sources already in 2009. Currently about 50 per cent of the Polish electricity from renewable source is produced from

biomass while a third comes from co-incinerating biomass in coal-fired power plants. Adding supports for the alternative technologies will lead to new opportunities to develop projects at industrial locations that produce large amounts of biomass.

This large share results in Poland's significant biomass potential and its large share of coal power plants that generate about 90 per cent of Poland's electricity. By using this existing infrastructure, Poland was able to significantly increase its share of renewable energies in just a few years. Today, 30 of the 39 Polish coal power plants are co-incinerating biomass.

The criticism of this practice, however, is immense. Most co-incinerating coal power plants do not use the waste heat (75 per cent). As a result of the boom in co-incineration in Poland, prices for energetically used biomass have increased considerably since 2006. This has led to an enormous increase of import of biomass, like wood pellets from Russia. The Polish authorities want to create a new solution considering the new law on renewable energies. The novelty is that, different types of electricity generation will receive different compensations. The compensation for smaller mono-incinerators should, in some cases, be more than four times as high as the compensation for large coal power plants that do not utilize their waste heat. At the same time, the growth of renewable energy should once again go faster. The new Polish law on renewable energy will be an opportunity for the development of new biomass projects, for instance using waste of the wood or furniture industry. As the current policy of co-incineration support is prevailing, such locations were not important. Only 11 biomass power plants are operational at the about 450 locations that produce large amounts of biomass. The largest plants are operational at the three large Polish pulp mills and they are also co-incinerators. The eight existing mono-incinerators at industrial locations are considerably smaller and they produce less than five per cent of the Polish biomass electricity. The new legislation is debated. The large state-owned energy providers especially oppose decreasing subsidies for co-incineration. This barrier, however, will have to end soon as it is urgent to change the situation. The delay of the new law has resulted in temporarily slumping prices for renewable energies certificates. New power plants that were constructed according to the new law should become operational. Nevertheless, old co-incinerators continue to produce, even though many of them should be closed due to the announced legislative changes. This results in overcapacities and declining prices that affect the power plant operators themselves. The law on renewable energies is currently envisaged to come into force.

The Polish market for renewable energies has reached a turning point. The new Polish Act on Renewable Energy Sources entails new opportunities for developing biomass power plant projects including locations close to the wood or furniture industry. At present, mono-incinerators at such locations produce less than five per cent of the Polish electricity from biomass. Most potentially favourable locations have not yet been utilised. Instead, almost 80 per cent of the biomass electricity and more than a third of the renewable energies in Poland are generated through co-incinerating biomass in coal power plants. The new Polish Act on Renewable Energy Sources is scheduled to come into effect in the second half of 2013. According to this law, the support of co-

incinerators should decrease significantly in the future.

By contrast, smaller biomass power plants, mono-incinerators and electricity generation by using combined heat and power (CHP) technology should receive stronger support. At the same time, the goals in terms of developing and boosting renewable energies are once again increasing. In light of this development, experts have jointly analysed the market for electricity generation from solid biomass in detail. The report focuses on the identification of industrial locations that produce large amounts of biomass.

	Dependent variables					
Independent variables	ecoinn. prod.	SE	ecoinn. process	SE	ecoinn. organis.	SE
Constant	-0,63	7,88	13,7	9,71	-1,32	11,37
High material prices	0,34*	0,1	0,42	0,22	0,31*	0,15
High energy prices			-0,71*	0,29		
Expected high energy prices.			0,48*	0,22		
R2	0,42		0,32		0,15	
F (p value)	9,86(<0,001)		3,61 (<0.02)		6,36(<0,04)	

Table 1 Regression results for introduction of product, process and organisational ecoinnovation in European countries

Despite Poland's positive developments, there is room for improving Poland's energy strategy. First, a more integrated energy and climate policy is needed to put Poland firmly on a low-carbon path while enhancing energy security. Second, energy policy could put more emphasis on promoting competition to make the energy markets more efficient. Decarbonising Poland's power sector will be a particularly significant challenge requiring huge investments. Coal accounts for 55% of Polish primary energy supply and 92% of electricity generation, raising significant climate change and environmental challenges. To this end, Poland's efforts to improve energy efficiency and to diversify the country's energy mix are praiseworthy and should be pursued. The government's attention to R&D on clean coal technologies, including carbon capture and storage (CCS) is also encouraging. For further reading on the Polish energy situation, please refer to [Voigt et al. Enviroinfo 2014].

4 Analyses

In order to identify the factors influencing on introducing in companies eco-innovations in Europe, a multiple regression model was applied, and European EU enterprises data were analysed. In the model ecoinnovativeness measured as a proportion of enterprises which implemented particular kinds of environmentally-friendly innovations in 2009–2010 was chosen as a dependent variable. Independent variables used in models were factors influencing the decisions by companies as to whether they should implement eco-innovations. For the given objects of our analysis, i.e. the 27 groups of enterprises from EU countries, the values of independent variables were calculated as weighted means, where the weights were fractions of inquired respondents assessing given factors as very important and somewhat important.

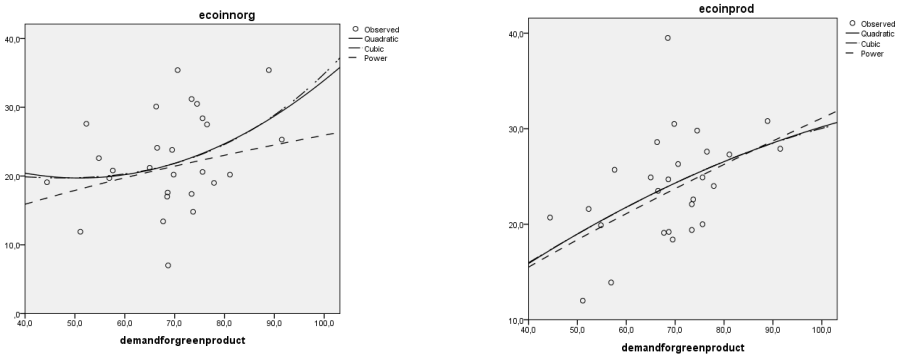


Fig. 1 Introduction of product ecoinnovation as a function of demand of green product on the market-approximation (left) Introduction of the process ecoinnovation as result of demand of green product on the market (right)

According to the results of the calculations, the only factor which positively influenced product eco-innovation implementation appeared to be an increase in high material prices (table 1). With respect to process eco-innovations it may be concluded that the current high prices of energy used in production processes and expected high energy price were only significant variables, and higher energy prices acted contrary to our theory-based expectations as a sedative to the implementation of new or significantly improved eco-innovative production processes or new methods.. Considering organizational ecoinnovation, the only significant variable was high material prices. Other factors were not significant, therefore they were not included in the model. We received positive correlation among demand for green product, process organisational ecoinnovation.

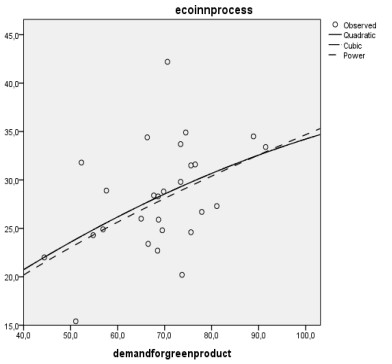


Fig. 2 Introduction of the organisational ecoinnovation as result of demand of green product on the market-approximation

7 Conclusions

There are considerable concerns related to the development of eco-innovativeness in some European countries including Poland. First, politicians do not recognize the significance of eco-innovations, or generally innovations. Strong lobby supports obsolete industries and decision makers seem to be not interested enough in development of eco-innovations. On the one hand, there is no pressure on research in new eco-innovative solutions. On the other hand, academic and R&D centers are unable to satisfy the demands of the industry. Eco-innovations are perceived as the driving factor of the third transformation in Poland and other EU members (following system transformation and the EU accession). The state should play the key role in this process – on the one hand it should inspire the demand for eco-innovations and on the other it should assure conditions for increasing interest in such solutions. Despite significant progress, the Polish economy has low productivity and high GHG emission intensity. Poland still has a lot to do to become an economy of high material and energy efficiency. This includes development of a necessary legal and institutional background. Such a transformation also requires fundamental changes in education and the behavioral patterns of citizens and companies so that we become a society of sustainable material consumption and move toward a green economy, where development is decoupled from material and energy use.

Among the factors very strongly stimulating eco-innovation implementation the respondents listed expected future increases in energy prices, current high energy prices and current high materials prices. Cooperation between research institution and companies is not satisfactory in some European countries. Poland stringent environmental regulation as main driver of eco-innovation. Those countries experience limited interest, lack of collaboration between business and research institutions, and limited budget for research at this area.

Literature

[CHdR00] Carrillo-Hermosilla, J., del Rio, P.: Diversity of eco-innovations: reflections from selected case studies. Journal of Cleaner Production 18, p. 1073-1083.

[At11] Attitudes of European entrepreneurs towards eco-innovation Summary Fieldwork: Flash EB No 315 2011

[CF99] Cleff, T. and Rennings, K. Determinants of environmental product and process innovation. Eur. Env., 9: 191–201, 1999.

[Ci10] Community Innovation Survey/CIS/ Eurostat, 2010. http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/cis

[GP09] Del Rio Gonzalez P.: The empirical analysis of the determinants for environmental technological change: A research agenda", Ecological Economics , 68, p. 861 – 878, 2009.

[Eco08] ECO-DRIVE: A framework for measuring eco-innovation:typology of indicators based on causal chains, Final Report, 2008m, FP6-2005-SSP-5-A. Retrieved from http://www.eco-innovation.eu/

[ECO12] Eco-innovation observatory, Methodological Report, 2012

[Eco12] Eco-innovation in Poland: 2012 update,

[Eco12a] Ekoinnowacje, to korzyści dla gospodarki i środowiska. Retrieved from zielondziennik.pl, 2012

[FJ96] Fussler, C. & P. James, 1996; Driving Eco-Innovation: A Breakthrough Discipline for Innovation and Sustainability, Pitman Publishing: London, 364 p.1996

EN21 Final Energy Consumption Intensity http://www.eea.europa.eu/data-and-maps/indicators/en21-final-energy-consumption-intensity/en21-final-energy-consumption-intensity

[HRR12] Horbach J, Rammer C., Rennings K.: Determinants of eco-innovations by type of environmental impact — The role of regulatory push/pull, technology push and market pull, Ecological Economics, pp. 112-122.2012

[FHR07] Frondel M., J. Horbach*and K. Rennings, End-of-Pipe or Cleaner Production? An Empirical Comparison of Environmental Innovation Decisions Across OECD Countries, Business Strategy and the Environment, 16, 571–584 (2007)

[GM i94] Green, K., McMeekin, A., Irwin, A., Technological trajectories and R & D for environmental innovation in UK firms. Futures 26, pp. 1047–1059,.1994

[KK11] Kassenberg A., P. Kassenberg Ecoinnovation in Poland, Ecoinnovation Observatory, 2011 Country Profile 2011: Poland

[Ka09] Kammerer , D. The effects of customer benefit and regulation on environmental product innovation. Empirical evidence from appliance manufacturers in Germany, Ecological Economics , Vol. 68, p. 2285-2295

[KRP08] Kemp , R ., P. Pearson: Final report MEI project about measuring eco - innovation, Maastricht, www.merit.unu.edu \ MEI . 2008

[Po06] Popp, D.: International Innovation and Diffusion of Air Pollution Control Technologies: The Effects of NO x and SO 2 Regulation in the US, Japan, and Germany, Journal of Environmental Economics and Management , 51/1 No. 1, 46 – 71, 2006

[RRZ07] Rehfeld K.; K. Rennings and A. Ziegler (2007), "Determinants of Environmental Product Innovations and the Role of Integrated Product Policy – An Empirical Analysis", Ecological Economics , Vol. 61, pp. 91 - 100.2007

[RM08] Reid A., Miedzinski M. (2008), Eco-innovation final report for sectoral innovation watch, www.technopolis-group.com

[Wo99] Wozniak, L. Mozliwe kierunki rozwoju ekoinnowacyjności . in: Ekoinnowacyjnosc dzis i jutro-wyzwania, bariery rozwoju oraz instumenty wsparcia. PARP, Warsaw1999

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Digitalisierung und Informationssysteme der Nachhaltigkeitsberichterstattung - Wie Institutionen IKT und Software-Werkzeuge einsetzen, um Kommunikation und Nachhaltigkeitsleistung zu verbessern

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Digitalisierung und Informationssysteme der Nachhaltigkeitsberichterstattung - Wie Institutionen IKT und Software-Werkzeuge einsetzen, um Kommunikation und Nachhaltigkeitsleistung zu verbessern

Prof. Dr. Ralf Isenmann[1], Prof. Dr. Jorge Marx Gómez[2]

Abstract: Der Workshop „Digitalisierung und Informationssysteme der Nachhaltigkeitsberichterstattung - Wie Institutionen IKT und Software-Werkzeuge einsetzen, um Kommunikation und Nachhaltigkeitsleistung zu verbessern“ verfolgt das Ziel, die Digitalisierung und den Einsatz von Informationssystemen als Herausforderung, Treiber und Erfolgsfaktor für die Nachhaltigkeitsberichterstattung von Unternehmen, Hochschulen und anderen berichterstattenden Akteuren zu untersuchen. Den verschiedenen, an der Nachhaltigkeitsberichterstattung Beteiligten sollen eine substanzielle Orientierung und robuste Handlungsempfehlungen angeboten werden, um die sich ergebenden Gestaltungschancen ausschöpfen zu können.

Der Workshop und seine Beiträge greifen aktuelle Facetten des Rahmenthemas auf, darunter: Visualisierung der Nachhaltigkeitsleistung, Darstellung von Nachhaltigkeit durch hybride Wertschöpfung, Nutzung von KPIs im Incident Management, Nutzung von Enterprise-Architecture-Ansätzen als Werkzeuge zur Nachhaltigkeitssteuerung, Nachhaltigkeitsberichterstattung von Hochschulen in Deutschland und Nachhaltigkeitskommunikation im eRecruiting. Insgesamt werfen die Beiträge jeweils ein Licht auf aktuelle Aspekte, die alle für die Digitalisierung und den Einsatz von Informationssystemen in der Nachhaltigkeitsberichterstattung einschlägig sind. Sie helfen Institutionen, die Gestaltungschancen moderner IKT und Software-Werkzeuge noch besser einzusetzen, um ihre Kommunikationsaktivitäten sowie vor allem ihre Nachhaltigkeitsleistung weiter zu verbessern.

Den Mitgliedern des Programmkomitees danken wir sehr für ihre Expertise bei der Ausgestaltung des Workshop-Programms sowie bei der fachlichen Begutachtung der eingereichten Beiträge:

- Prof. Dr. Hans-Knud Arndt, Otto-von-Guericke-Universität Magdeburg,
- Dr. Remmer Sassen, Universität Hamburg,
- Prof. Dr. Stefan Schaltegger, Leuphana-Universität Lüneburg,
- Prof. Dr. Frank Teuteberg, Universität Osnabrück.

[1] Hochschule München, Nachhaltiges Zukunftsmanagement, Am Stadtpark 20, 81243 München, ralf.isenmann@hm.edu
[2] Universität Oldenburg, Wirtschaftsinformatik, Ammerländer Heerstr. 114-118, 26125 Oldenburg, jorge.marx.gomez@uni-oldenburg.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

A new Visualization Approach for Sustainability

Matthias Mokosch[1], Torsten Urban[2], Hans-Knud Arndt[3]

Abstract: Organizations sustainability is increasingly measured by the general public. But does a suitable way always exist to present both the target sustainability as well as the current sustainability in a stakeholder oriented manner? Today especially the management needs a tool, which is efficient and easy to understand to reach a stakeholder oriented sustainability presentation. In our contribution we discuss advantages and disadvantages of existing approaches to visualize sustainability. Moreover, we present our triangular pyramid of sustainability, which is the result of a new three-dimensional visualization approach. It solves many problems of existing sustainability visualization approaches. In a nutshell, our contribution presents a visualization approach that enables manager to determine the quality and quantity of sustainability as well as to conduct sustainability comparisons between and within organizations but also between different periods

Keywords: Triangle of Sustainability, Triple Bottom Line, Visualization, Grand Management Information Design

1 Introduction

Organizations are increasingly measured by means of their sustainability and use the positive associations of the stakeholder for both as marketing tool but also as instrument to present their products in a positive way. Due to the numerousness influencing factors and potential key performance indicators it is not easy to present sustainability results in a stakeholder oriented manner. Especially, the strategic management must be able to get crucial sustainability information as well as to generate ad hoc reports. This is a very important point in terms of competitiveness. A visualization of both the influencing factors but also the existing sustainability would be preferable. Despite, existing approaches fulfil these expectations insufficiently. Therefore, the management needs a tool that is efficient and easy to use to be able to make high quality decisions on strategic level concerning sustainability. Within the scope of basic research our contribution presents a graphical approach to close the mentioned gap above. Therefore, we introduce existing approaches and discuss their advantages and disadvantages. The outcome of this is that the visualization of all three pillars of sustainability in a two dimensional room is strongly

[1] Otto-von-Guericke-Universität Magdeburg, Institut für Technische und Betriebliche Informationssysteme -Managementinformationssysteme-, Universitätsplatz 2, 39106 Magdeburg, matthias.mokosch@ovgu.de
[2] Städtische Werke Magdeburg GmbH & Co. KG, Am Alten Theater 1; 39104 Magdeburg, torsten.urban@sw-magdeburg.de
[3] Otto-von-Guericke-Universität Magdeburg, Institut für Technische und Betriebliche Informationssysteme -Managementinformationssysteme-, Universitätsplatz 2, 39106 Magdeburg, hans-knud.arndt@iti.cs.uni-magdeburg.de

connected with interpretation or visualization problems. As solution, we shift the visualization problem in a three-dimensional space. As a result, our approach eliminates the disadvantages of existing approaches. The way to visualize sustainability as triangular pyramid keeps the similarity to both the triangle of sustainability but also to the Triple Bottom Line.
With our approach, the management gets a visualization method which is similar to existing approaches and able to visualize sustainability quantitatively and qualitatively. Furthermore, our approach supports manager to conduct sustainability comparisons between and within organizations but also between different periods.

2 Analysis of Existing Approaches

2.1 Theoretical Background

The term "sustainable development" or shorter "sustainability" is defined in the report "Our Common Future" as "Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs." [Un 87]. This abstract definition was extended about the three pillars of sustainability environmental, economic and social on Rio Conference in 1992. In 1998, the extended definition was mentioned in the final report of the committee of inquiry [De 98]. The Global Reporting Initiative (GRI) provides an interpretation of the content of the three pillars of sustainability in terms of a guideline to generate standardized sustainability reports. "Environmental" describes the impact on living and non-living natural systems and includes the elements air, earth and water. Unlike environmental, "economic" deals with impacts that influence economic conditions on local, national or international level. "Social" includes labor practices, human rights, society and product responsibility [Gl 13]. The extension of the sustainability definition mentioned above is also known as Triple Bottom Line (TBL) [Pu 12] and led to different graphical representations. In the following, we introduce some of these visualizations.

2.2 Approaches to Visualize Sustainability

The traditional representation of a sustainability triangle is an equilateral triangle as shown in figure 1. Each vertex represents one of the three sustainability aspects. The identical length of every edge shows the equal importance of all sustainability aspects. [In 14]

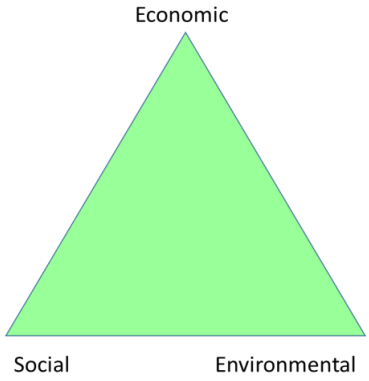


Fig. 1: The traditional sustainability triangle (adapted from [Aa 14]).

A similar representation of the sustainability triangle shows figure 2. Unlike the traditional sustainability triangle, the sustainability perspectives in figure 2 are not represented by the vertexes but by the sides of the triangle. Inside the triangle the sustainability perspectives get mixed. This illustrates that the sustainability perspective influence each other [Kl 09].

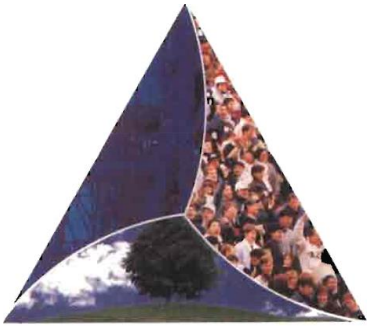


Fig. 2: The sustainability triangle of committee of inquiry [De 98].

The representation in figure 3 is an extension of the triangle approach. It is obviously that sustainability lean on three pillars: Social, economic and environmental. This static representation emphasizes the necessity of all pillars to avoid any tilt of sustainability.

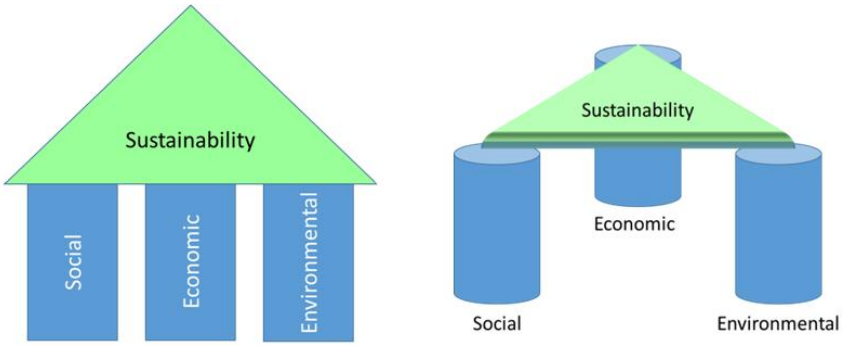


Fig. 3: The traditional visualization of sustainability in terms of pillars (on the left) and its representation in a three dimensional space (on the right) (adapted from [Kl 09]).

Another not triangle based representation of sustainability is shown in figure 4 in terms of an intersection model. Each circle represents a sustainability perspective. The intersections generated by at least two circles illustrate the trade-off between specific sustainability perspectives [Kl 09]. In a literal sense the sustainability is just located in the middle of the illustration and is generated by the intersection of all three sustainability aspects [OE 00].

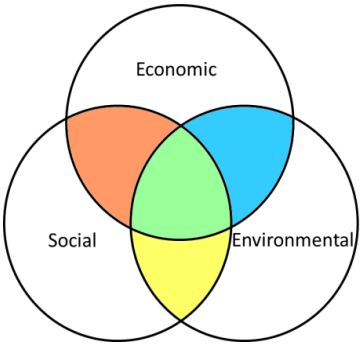


Fig. 4: The intersection model of sustainability (adapted from [Kl 09]).

2.3 Domain of Sustainability Visualization

The triangle of sustainability can be used to position business lines of action and different indicators along the triangle sides depending on their relevance [N. 98]. Adding additional information on the triangle sides is an extension of this triangle approach. Every triangle side is with regards to the content divided into both an efficiency area and an effectiveness area. This partition supports the discussion about the positioning of different targets on the

triangle sides. Within this approach, the vertexes represent the effectiveness targets economic, environmental and social. [DH 02]
An approach using the triangle area is the fractal triangle. The idea of this approach is that both the acting of humans as well as their initialized mass flow should be oriented on the example of the nature. The meaning and interpretation of the triangles created by fractal division is determined by the two shortest distances of the triangle to the specific vertexes. This approach allows the classification of sustainability questions within the triangle [Kl 09].

The triangle of Gibbs as shown in figure 5 is a mathematical approach which makes it possible to calculate a concrete value for each sustainability perspective. Identically to some approaches mentioned before, the vertexes in the triangle of Gibbs represent the three sustainability perspectives. Each triangle side has the function of a scale from 0 to 100 percent. For each point within the triangle the specific sustainability values can be calculated using parallel translation. A characteristic of this approach is that adding the three sustainability values always leads to 100 percent. An implication is the following formula:

$$x + y + z = 100\%, \text{with } x, y, z \in [0\%, 100\%] \quad (1)$$

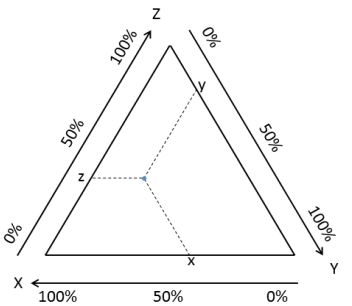


Fig. 5: The triangle of Gibbs (adapted from [Kl 09]).

3 Problems

In the following, we like to introduce visualization problems connected with the mentioned visualization approaches as well as with the triangle of Gibbs. In order to reach sustainability both representations, the traditional sustainability triangle and the committee of inquiry, put the focus on the visualization of the equal importance and on the corporation of all sustainability perspectives. Changes in one or more sustainability perspectives as well as their impacts cannot be visualized because of the triangle sides which are defined as equilateral. Figure 3 shows the sustainability perspectives as

fundament of the sustainability above. If the middle pillar or the outer pillars in figure 3 are removed, the sustainability will be stable anymore. This is a contradiction to the idea that all perspectives must be fulfilled to reach sustainability at all. This problem is eliminated in the three-dimensional representation but creates the impression that all sustainability perspectives must have identical values to reach sustainability. An improvement of one sustainability perspective leads to a higher pillar and the outcome of this would be a tilt.
Within intersection model the sustainability perspectives are visualized by three overlapping circles. Unlike the pillar or triangle representations, the intersection model allows to assign sustainability aspects to more than one sustainability perspective. A problem is to find the specific degree of intersection. On the one hand, too little intersection creates both the impression of small sustainability importance as well as increasingly independent and isolated sustainability perspectives. On the other hand, too strong interaction leads to sustainability perspectives which are not or only weak distinguished from each other [Kl 09].

This shortcoming is eliminated at the triangle of Gibbs. All points within and on the triangle sides as well as the vertexes can be visualized as components of the specific sustainability perspectives. According to the definition (equation 1), the sum of all sustainability perspectives is 100 percent. This is a disadvantage for interpretation. This approach allows the segmentation of one sustainability value into three values, but only as relative and not absolute values, one for each sustainability perspective. This limitation leads to visualization problems, because a comparison between absolute values is not possible.

It has to be mentioned that the visualization approaches explained above have different focuses. Moreover, each of the introduced approaches comes along with disadvantages as well as inherent risks of misinterpretation. This is mostly caused by the attempt to visualize all three sustainability perspectives within a two-dimensional space. From a mathematical point of view the third sustainability aspect cannot expressed independently from the other one. The third aspect is always determined by the other one. Consequently, our contribution aims to develop a visualization approach which supports:

- First, a representation of changes in one or more sustainability perspectives.
- Second, a visualization of sustainability as result of three sustainability perspectives.
- Third, the presentability of sustainability which takes dependencies between sustainability aspects into account.
- Fourth, a calculation of one sustainability value.
- Fifth, the presentability of equal importance of all sustainability perspectives.

4 Contribution

According to TBL, sustainability can be seen as a function, which has three variables. These are the pillars of sustainability. Equation 2 shows this relationship, whereas S_{result} expresses the resulting sustainability. x_{social}, $y_{environmental}$, and $z_{economic}$ stand for the concrete values of the social, environmental and economic pillar of sustainability.

$$S_{result} = f(x_{social}, y_{environmental}, z_{economic}) \tag{2}$$

As we have shown in the last chapter, existing approaches try to visualize three variables in a two-dimensional space. The resulting problem is that at least one variable dependent on the other two. In our contribution, we present a solution transferring the visualization in a three-dimensional space.

4.1 Assumptions

In order to enable the transfer of the visualization into a three-dimensional space as well as to limit the following explanations on the graphical representation, we have to make two assumptions.

First, we assume that it is possible to calculate a cumulative and normalized key figure for each pillar of sustainability. Therefore, it is possible to calculate a value between 0 and 100 percent for the social, environmental as well as for the economic pillar. This assumption is in line with the triangle of Gibbs, where values are in a range between 0 and 100 percent, too [Kl 09]. For example, we can determine the value of the overall key figure for the social pillar (x_{social}) summing up the partial results of the subdomains $i = 0 \rightarrow n$ which are a product of an influencing factor (k_i) and a quotient of the current value (x_i) and its maximum value (x_i^{max}), look at equation 3. The key values for the environmental ($y_{environmental}$), and economic pillar ($z_{economic}$), can be determined in the same way. In order to identify the relevant subdomains for each pillar, we suggest the GRI-Guidelines [Gl 13] as starting point. In organizations the required values can be delivered potentially by departments which are responsible for finance, controlling, social responsibility and environment. However, a detailed discussion about this is not part of our contribution.

$$x_{social} = \sum_{i=0}^{n} k_i \times \frac{x_i}{x_i^{max}} \tag{3}$$

Second, we assume that the pillars of sustainability are independent form each other. Hence, changing one pillar of sustainability has no impacts on the other pillars. For example, restructuring a production process as part of a process improvement could increase the output but also the operational result without influencing both the social and the environmental pillar of sustainability. In addition, a new environmental protection agenda could probably be influence the economical pillar, but it is not determined whether it will be positive or negative. The operational result can decrease because of higher costs

or increase because of a better image and the resulting sales quantity.

4.2 Visualization

Considering our assumptions, we can map the three pillars of sustainability into an orthogonally, three-dimensional coordinate system. The axis X_{social} corresponds to the social, $Y_{environmental}$ corresponds to the environmental, and $Z_{economic}$ corresponds to the economical pillar of sustainability. We can identify the point $P^S\langle x_{social}|y_{environmental}|z_{economic}\rangle$ plotting the key figure values of sustainability onto the corresponding axes. However, for a better compatibility with existing approaches, we do not take P^S into account. In our contribution, we look at the three points $P^X\langle x_{social}|0|0\rangle$, $P^Y\langle 0|y_{environmental}|0\rangle$, and $P^Z\langle 0|0|z_{economic}\rangle$ which are located on the coordinate axes. Connecting these points, leads to the three-dimensional triangle T^{XYZ}, which looks familiar to other visualization approaches, look at figure 6.

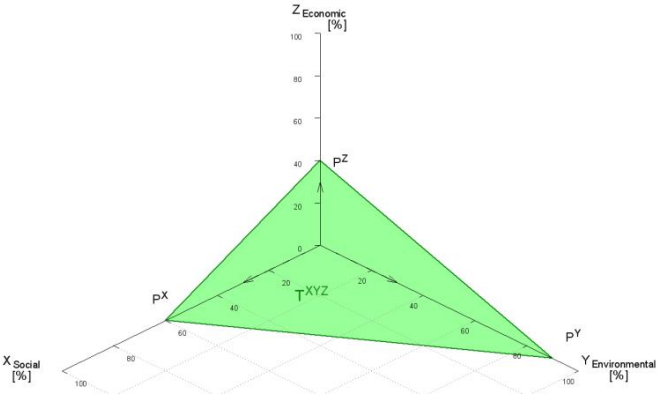


Fig. 6: Three-dimensional triangle of sustainability.

The triangle T^{XYZ} visualizes the cooperation between the three pillars of sustainability. The surface area of this triangle depends directly on the values of the corresponding key figures. If one key figure value increases, the surface area of the triangle increases, too. Furthermore, it is possible to draw three more triangles considering the origin of coordinates $P^0\langle 0|0|0\rangle$. These triangles are T^{XY0} with the points P^X, P^Y, and P^0, T^{X0Z} with the points P^X, P^0, and P^Z, and T^{0YZ} with the points P^0, P^Y, and P^Z, look at figure 7.

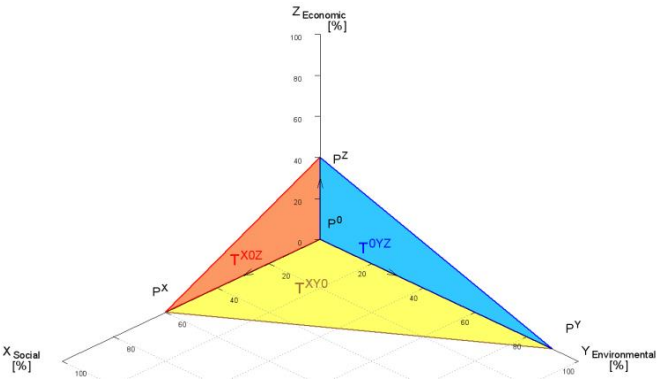


Fig. 7: Triangles of the cooperation between the pillars of sustainability.

The triangle T^{XY0} represents the cooperation between the social and environmental pillar of sustainability and is independent from the economic pillar. In the same way, triangle T^{X0Z} represents the cooperation between the social and economic pillar of sustainability and is independent from the environmental pillar, and triangle T^{0YZ} represents the cooperation between the environmental and economic pillar of sustainability and is independent from the social pillar.

We get a three-dimensional field, a triangular pyramid TP^N consisting of the points P^X, P^Y, P^Z, and P^0 by connecting these triangles. We call it triangular pyramid of sustainability.

5 Special Cases of the Visualization

In this chapter, we show the characteristics of triangular pyramid of sustainability TP^N. Our analysis turned out five special cases:

- First, exactly one key figure value of the pillars of sustainability becomes zero. In this case, the triangular pyramid of sustainability "collapses" to a two-dimensional triangle, which is spanned by two coordinate axes. If $x_{social} = 0$ the resulting triangle will be T^{0YZ}, if $y_{environmental} = 0$ the resulting triangle will be T^{X0Z}, and if $z_{economic} = 0$ the resulting triangle will be T^{XY0}.
- Second, exactly two key figure values of the pillars of sustainability become zero. Under this term, the triangular pyramid of sustainability "collapses" to a segment between the origin of coordinates P^0 and point P^S. It applies: $y_{environmental} = z_{economic} = 0 \rightarrow P^S = P^X$, $x_{social} = z_{economic} = 0 \rightarrow P^S = P^Y$, and $x_{social} = y_{environmental} = 0 \rightarrow P^S = P^Z$.

- Third, all three key figure values of the pillars of sustainability become zero. In this special case, the triangular pyramid of sustainability is equal to the origin of coordinates P^0.

However, in all three cases mentioned above we have to avoid using the word sustainability, because the missing of one or more pillars is a contradiction to the basic assumptions of the TBL.

- Fourth, all three key figures of the pillars of sustainability have the same value (unlike zero). In this term, the triangle T^{XYZ} become an equilateral triangle and equals the equilateral triangle of sustainability, look at figure 8.

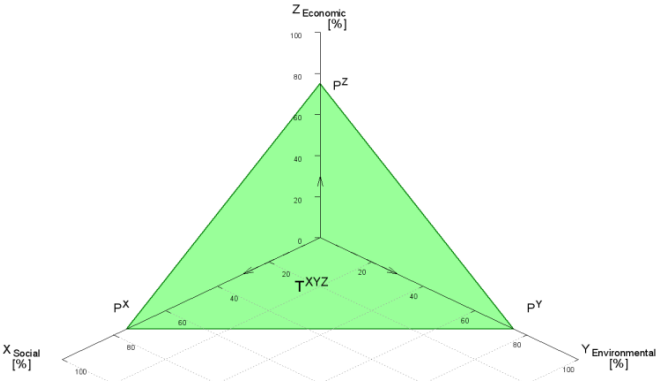


Fig. 8: Three-dimensional equilateral triangle of sustainability.

- Fifth, all three key figures of the pillars of sustainability have the value 100 percent. When all pillars of sustainability reach their maximum key figure value, $x_{social} = y_{environmental} = z_{economic} = 100\%$, the spread of the triangular pyramid of sustainability is as high as possible. This case describes an ideal situation. We call the resulting pyramid the ideal triangular pyramid of sustainability with equilateral base area TP^{N*}, look at figure 9.

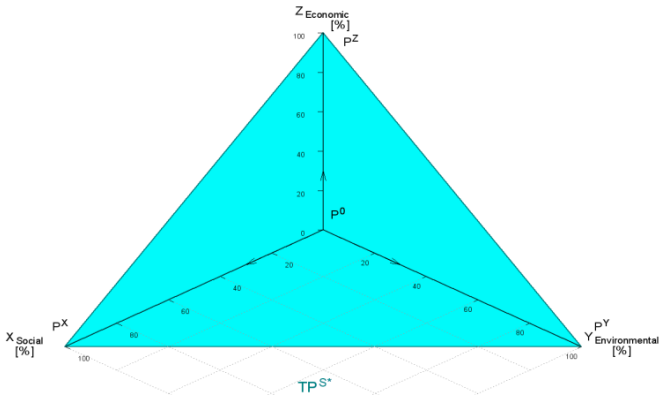


Fig. 9: Ideal triangular pyramid of sustainability with equilateral base area.

6 Discussion

In chapter two, we present several visualization approaches of sustainability and discuss their pro and cons. Now, we compare our triangular pyramid of sustainability with these.

It is impossible to visualize the change of one or more pillars of sustainability with the classic triangle of sustainability or the triangle of the committee of inquiry, because of equilateral triangles. In contrast, our triangular pyramid of sustainability allows this behavior (requirement 1). In addition, an equality of all pillars can be visualized, like described in special case 4. Moreover, our approach also emphasizes the equal importance of all pillars (requirement 5) by defining an optimum, look at special case 5.

The traditional pillar model of sustainability suggests that sustainability is in part independent of the pillars, like a roof of a house. If one or two pillars are missing, the roof will still exist. Our approach defines a strict relationship between the three pillars of sustainability and the sustainability itself qualitatively and quantitatively (requirement 3), look at special case 1 up to 3. In addition, the problem of a tilt by different key figures values is also eliminated without ignoring the dependence between the pillars and the overall sustainability.

According to the intersection model, figure 4, our triangular pyramid of sustainability also indicates the intersections between the pillars of sustainability with the triangles T^{XY0}, T^{X0Z}, and T^{0YZ} and emphasizes the cooperation of the pillars (requirement 2). Furthermore, our triangular pyramid of sustainability solves the problem of the undefined rate of intersections. The area of our triangular pyramid of sustainability can be calculated with the key figure values of the pillars of sustainability (requirement 4).

The main criticism point of visualizing the sustainability with the triangle of Gibbs is that

all points at the triangle of Gibbs always sum up to 100 percent. This fact suggests that the sustainability is always at 100 percent, too. The triangle of Gibbs allows assessing the quantity of the pillars of sustainability, but a comparison between concrete values of sustainability is impossible. Our approach enables to quantizing the sustainability using basic mathematical knowledge. Therefore, it is possible to make comparisons between different triangular pyramids of sustainability as well as the ideal triangular pyramid of sustainability, under the condition of equal influencing factors of all subdomains.

Although, our approach has some improvements in comparison to existing methods, but we also have to mention, that there is no software implementation, yet. We describe the theoretical background of our visualization approach. A quick solution could be realized with Excel. A more complex implementation should integrate interfaces to data suppliers, reporting tools as well as other application systems. This practical aspect is left for future work.

7 Conclusion and Future Work

First, we clarify the problem and convey the theoretical basics. Next, we present existing methods to visualize sustainability as well as their interpretations. Furthermore, we analyze the existing methods regarding their pros and cons and present a list of requirements based on our findings. In the main part of our contribution, we develop an own visualization approach, which allows to display the sustainability in the form of a triangular pyramid in a three-dimensional space. We verify our approach against some special visualization cases as well as against our list of requirements.
In conclusion, we developed a management tool, which fulfil the requirements mentioned in chapter one. This tool allows a qualitative and quantitative visualization of sustainability and supports the management to conduct sustainability comparisons between and within organizations but also between different periods. Our triangular pyramid of sustainability favors existing approaches, like the triangle of sustainability.

Future work has to deal with developing an efficient application of our approach. Determining a concrete key figure of sustainability S_{result} is also left for future. For example, both variations, the area of the triangle T^{XYZ}, as well as the volume of the pyramid TP^{N} are conceivable. Furthermore, the operationalization of the key figures x_{social}, $y_{environmental}$, and $z_{economic}$ for the pillars of sustainability, and defining a function method $f(x_{social}, y_{environmental}, z_{economic})$ are topics for future work. In addition, we prefer a more general discussion about our assumptions as well as the usability of three-dimensional figures.

References

[Aa 14] Aachener Stiftung Kathy Beys: Drei Säulen Modell. http://www.nachhaltigkeit.info/artikel/1_3_a_drei_saeulen_modell_1531.htm, 23.07.2014.

[De 98] Abschlußbericht. Konzept Nachhaltigkeit Vom Leitbild zur Umsetzung, 1998.

[DH 02] Dyllick, T.; Hockerts, K.: Beyond the business case for corporate sustainability. In Business Strategy and the Environment, 2002, 11; S. 130–141.

[Gl 13] Global Reporting Initiative: G4 - Sustainability Reporting Guidelines. Reporting Principles and Standard Disclosures. https://www.globalreporting.org/resourcelibrary/GRIG4-Part1-Reporting-Principles-and-Standard-Disclosures.pdf, 23.07.2014.

[In 14] Institute Construction and Environment: Triangle of sustainability. http://construction-environment.com/hp6259/Triangle-of-sustainability.htm, 09.04.2014.

[Kl 09] Kleine, A.: Operationalisierung einer Nachhaltigkeitsstrategie. Ökologie, Ökonomie und Soziales integrieren. Gabler, Wiesbaden, 2009.

[N. 98] N.N.: Wirtschaftliche Leistungsfähigkeit, sozialer Zusammenhalt und ökologische Nachhaltigkeit. Drei Ziele - ein Weg. J.H.W. Dietz, Bonn, 1998.

[OE 00] Towards Sustainable Development. Indicators to measure progress. OECD Publishing; OECD, Paris, 2000.

[Pu 12] Pufé, I.: Nachhaltigkeit. UTB, 2012.

[Un 87] United Nations: Report of the World Commission on Environment and Development: Our Common Future. Development and International Co-operation: Environment. http://www.un-documents.net/wced-ocf.htm, 17.03.2014.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Nachhaltigkeit durch Hybride Wertschöpfung – Entwicklung eines Reifegradmodells

Christina Niemöller[1], Niclas Bärtling[2] und Oliver Thomas[3]

Abstract: Die Nachhaltigkeit von Unternehmen, Produkten und Dienstleistungen kann durch Hybride Wertschöpfung gesteigert werden. Dennoch existiert bisher kein umfassendes Modell, mit dem der Reifegrad der Nachhaltigkeit durch hybride Lösungen bestimmt und berichtet werden kann. Deshalb soll in diesem Beitrag aufbauend auf dem State-of-the-Art der Nachhaltigkeit durch Hybride Wertschöpfung ein Reifegradmodell entwickelt werden, welches sich u.a. aus einem Katalog an Ausprägungen für die Dimensionen der Nachhaltigkeit Ökonomie, Ökologie und Gesellschaft sowie deren Schnittmengen zusammensetzt. Zudem werden Merkmale für das Erreichen einer bestimmten Reifegradstufe herausgearbeitet und in einem Demonstrationsbeispiel evaluiert. Somit soll eine Aussage möglich werden, inwiefern (a) die Hybride Wertschöpfungsaktivität mit den Aktivitäten im Bereich Nachhaltigkeit übereinstimmen und (b) welche Auswirkungen die Implementierung einzelner hybrider Maßnahmen auf die Nachhaltigkeit haben. Der vorliegende Beitrag liefert ein Modell, das als Ausgangspunkt für weitere empirische Forschung und die prototypische Implementierung verwendet werden kann.

Keywords: Hybride Wertschöpfung, Product-Service Systems, Nachhaltigkeit, Reifegradmodell

1 Einleitung

Dass ökonomisches Wachstum mit ökologischem und sozialem Druck einhergeht, wurde bereits vielfach in der Wissenschaft und Praxis diskutiert [CL09]. Es stellt sich die Frage, wie man den Gedanken der Nachhaltigkeit erfolgreich in die Unternehmenspraxis implementieren kann. Als einer der ersten Ansätze für eine nachhaltigere Entwicklung wird die sogenannte Hybride Wertschöpfung angeführt [Ba07, CL09]. Als erstes entwickeltes Konzept für dieses neue Wertschöpfungsverständnis wurde in der Fachliteratur der Beitrag von GOEDKOPP ET AL. identifiziert [Ba07, Go99, Pi10]. Dabei geht es neben dem Konzept der Nachhaltigkeit auch darum, Lösungen für die stetig komplexer werdenden Produkte und anspruchsvolleren Kunden zu entwickeln. Daher ist der Einsatz der Informationstechnologie (IT) notwendig, um eine Brücke zwischen Sach- und Dienstleistung zu schlagen [BT13]. Diese Kombination aus Produkt, Dienstleistung und der IT wird als Hybride Wertschöpfung oder Product-Service Systems (PSS) bezeichnet. Sie erhält auch bei der Entwicklung von neuen, nachhaltigeren Produkten eine immer größere Bedeutung

[1]Universität Osnabrück , Fachgebiet Informationsmanagement und Wirtschaftsinformatik, Katharinenstraße 3, 49074 Osnabrück, christina.niemoeller@uni-osnabrueck.de
[2] Universität Osnabrück , Fachgebiet Informationsmanagement und Wirtschaftsinformatik, Katharinenstraße 3, 49074 Osnabrück, nbaertling@uni-osnabrueck.de
[3] Universität Osnabrück , Fachgebiet Informationsmanagement und Wirtschaftsinformatik, Katharinenstraße 3, 49074 Osnabrück, oliver.thomas@uni-osnabrueck.de

in der heutigen Wirtschaft [NTF14, TLN10].

Es existieren bereits Ansätze in der Forschung, die den Einfluss von PSS auf die unterschiedlichen Dimensionen der Nachhaltigkeit zu beurteilen versuchen. Ein umfassendes Modell, welches die Effektivität der Umsetzung des Gedankens der Nachhaltigkeit in Unternehmen und den Einfluss von PSS auf diese misst, konnte jedoch noch nicht entwickelt werden. Deshalb stellt sich die Frage, wo genau PSS eingesetzt werden und wirken können, um die Nachhaltigkeit in der Unternehmenspraxis zu verbessern.

Das Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (2009) hat beispielsweise Empfehlungen für die Nachhaltigkeitsberichterstattung herausgegeben [Bu09]. Auf internationaler Ebene beschäftigt sich die Global Reporting Initiative (GRI) (2013) in ihren Richtlinien mit der Messung von Nachhaltigkeit [Gl13]. Des Weiteren werden die Nachhaltigkeitsberichte der deutschen Unternehmen in regelmäßigen Abständen vom Institut für ökologische Wirtschaftsforschung und dem future e.V. evaluiert. Auf diesen Evaluationen, den Richtlinien, den Nachhaltigkeitsberichten und dem State-of-Art von Nachhaltigkeit durch Hybride Wertschöpfung aufbauend soll untersucht werden, wie man messen und darstellen kann, wie stark die Hybride Wertschöpfung und die Nachhaltigkeit im Unternehmen ausgeprägt ist. Somit soll eine Aussage möglich werden, inwiefern (a) die Hybride Wertschöpfungsaktivität mit den Aktivitäten im Bereich Nachhaltigkeit übereinstimmen und (b) welche Auswirkungen die Implementierung einzelner hybridischer Maßnahmen auf die Nachhaltigkeit haben. Dazu wird ein Reifegradmodell entwickelt, dass zunächst auf die interne Kommunikation zur Effizienzsteigerung der eigenen Prozesse abzielt, aber im späteren Verlauf auch gemeinsame Stakeholder der hybriden Wertschöpfung (Produkthersteller, Operativer Dienstleister, IT-Dienstleister etc.) einbeziehen kann und somit zum „Involve me“-Berichterstattung werden [Sc14]. Aufbauend darauf können in weiteren Forschungsarbeiten der empirische Grad der Nachhaltigkeit von Unternehmen bestimmt, Gründe für die Abweichungen ermittelt und Maßnahmen zur Stärkung der Nachhaltigkeit entwickelt werden. Dazu ist im nächsten Schritt die prototypische Implementierung des vorgestellten Konzepts erforderlich.

2 Stand der Forschung

Zur Identifikation von verwandten Arbeiten und Aufdeckung der Forschungslücke wurde eine systematische Literatur Recherche nach vom Brocke [Br09] zum Thema *Messung und Darstellung von Nachhaltigkeit in der Hybriden Wertschöpfung durch Reifegradmodelle* durchgeführt.

In Anlehnung an [BT13, Br09, KPW08] wurden folgende Datenbanken durchsucht: ScienceDirect, ISI Web of Knowledge, Ebsco-Host, Springer Link, EmeraldInsight, Wiley online library, AIS electronic Library und IEEEXplore. Des Weiteren wurde parallel die Internetsuchmaschine Google Scholar durchsucht.

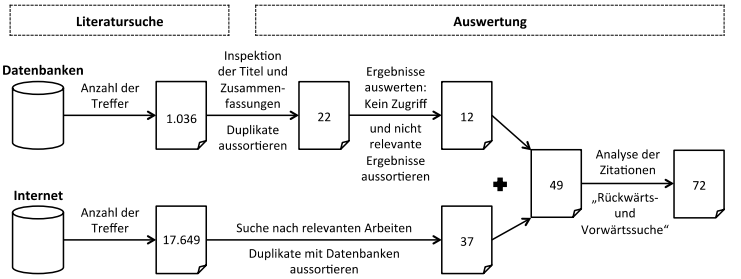


Abb. 1: Literaturrecherche im Überblick

Dabei wurden die folgenden Queries (deutsch und englisch) zur Kombination der Themenfelder *Nachhaltigkeit*, *Hybride Wertschöpfung* und *Reifegradmodelle* zur Suche genutzt: *[Nachhaltigkeit OR "nachhaltige Entwicklung" OR „Drei Säulen Modell") AND ("Hybride Wertschöpfung" OR PSS OR "Produkt Service" OR "hybrides Produkt" OR Kundenlösung) AND (Reifegradmodell OR Reifegrad OR Maturität OR messen OR Grad)]* sowie im Englischen *[sustainability OR "sustainable development" OR "triple bottom line") AND ("hybrid value creation" OR PSS OR "product service" OR "hybrid product" OR "customer solution") AND ("maturity model" OR maturity OR measure OR degree)].*

Zunächst ist festzuhalten, dass infolge der durchgeführten Literaturanalyse keine expliziten Arbeiten identifiziert werden konnten, welche sich mit der *Entwicklung eines Reifegradmodells zur Messung von Nachhaltigkeit in hybrider Wertschöpfung* beschäftigen. Aufgrund dessen werden in der inhaltlichen Literaturanalyse die Arbeiten berücksichtigt, die sich allgemein mit PSS und Nachhaltigkeit befassen und die Implikationen auf die Entwicklung des Reifegradmodells beinhalten.

Bereits 2000 beschreibt ROY das *Konzept nachhaltiger PSS*, welches sich aus der Entwicklung hin zu „cleaner production, eco-design and design for the environment" [Ro00] gebildet hat. Dabei werden die ersten Strategien beschrieben, welche verfolgt werden können, um einen geringeren negativen Einfluss auf die Umwelt zu nehmen. EPSTEIN UND ROY beschäftigen sich mit der Implementierung der drei Dimensionen der Nachhaltigkeit in das unternehmerische Handeln. Dabei präsentieren sie einen Rahmen, mit dem Treiber der Nachhaltigkeit festgelegt und gemessen werden können. Dabei ist es bedeutend, dass identifiziert werden muss, welche Produkte, Dienstleitungen und Prozesse in welchem Ausmaß langfristige Auswirkungen auf die Nachhaltigkeit haben [ER01].

Weiterführend bringt MONT in ihrem Artikel das *Potential von PSS* zur Sprache, die Auswirkungen von Produktion und Konsum auf die Umwelt zu minimieren. Des Weiteren wird festgehalten, dass die Charakteristika, welche PSS umweltfreundlicher machen sollen, zu diesem Zeitpunkt noch nicht erforscht waren [Mo02]. Darauf aufbauend untersucht TUKKER die verschiedenen Kombinationen aus Produkt und Dienstleistung und welchen zusätzlichen Nutzen diese auf die ökonomische und ökologische Komponente der Nachhaltigkeit haben. Dabei stellt sich heraus, dass die Effekte je nach Kombination durchaus

unterschiedlich ausfallen. Durch Leasingangebote können Nutzer von PSS sogar dazu verleitet werden, weniger verantwortungsbewusst zu handeln, weil sie nicht der Eigentümer des PSS sind. Dies könne zu negativen ökologischen Auswirkungen führen [Tu04]. Daran knüpfen PIGOSSO ET AL. an und beschäftigen sich mit der Frage, ob PSS zu mehr Nachhaltigkeit führt und was diese Entwicklung ausmacht. Dabei wird eine Auswahl der Treiber der Nachhaltigkeit aufgelistet. Außerdem wird beschrieben, dass Unternehmen bis dato PSS nicht der Umwelt wegen, sondern aus ökonomischen Gründen eingeführt hätten [Pi10]. Zu ähnlichen Schlüssen gelangen COLEN UND LAMBRECHT wobei sie Strategien beschreiben und erläutern, welche zu einer erfolgreichen Implementierung von nachhaltigen PSS führen sollen [CL10].

Da sich bis zu diesem Zeitpunkt die Forschungsbeiträge vor allem mit den Auswirkungen von PSS auf die ökonomische und ökologische Dimension der Nachhaltigkeit beschäftigt haben, widmen sich XING ET AL. dem Einfluss auf die Gesellschaft [XNL13, XWQ13]. Auch MYLAN beschäftigt sich mit dem Einfluss auf die Gesellschaft bzw. auf die Konsumenten von nachhaltigen PSS. Sie geht also im Vergleich zu anderen Arbeiten intensiv auf die Soziologie der Nachfrager von PSS ein. Dabei kommt sie zu dem Ergebnis, dass die Akzeptanz von PSS-Nutzern verbessert und die Bedürfnisse von Konsumenten nicht nur erfüllt, sondern gegebenenfalls auch transformiert werden [My14].

Mit den einzelnen Phasen des *Produktlebenszyklus* und deren Auswirkung vor allem auf die ökologische Dimension der Nachhaltigkeit in PSS beschäftigen sich u.a. FINNVEDEN ET AL. in ihrem Beitrag. Als relevanter Aspekt ist die Tatsache anzumerken, dass für die Messung von Auswirkungen auf die Nachhaltigkeit von Produkten und Dienstleistungen der gesamte Produktlebenszyklus betrachtet werden muss. Sonst könnten zum Beispiel negative Auswirkungen aus der Nutzungsphase in die Nachnutzungsphase verschoben werden, welche gar nicht erst betrachtet wird, sodass die PSS-Lösung nur vermeintlich nachhaltiger ist [Fi09].

Mit der *Messung von Nachhaltigkeit* beschäftigen sich u.a. HANSEN ET AL. sowie SCHMIDT ET AL., welche exemplarische Maßnahmen nennen, die zu mehr Nachhaltigkeit im IT-Management führen sollen [HGR09, Sc09a]. ALLEN HU ET AL. beschreiben verschiedene Kriterien, die zur Messung von nachhaltiger Performance in PSS genutzt werden können. Diese werden zwei Hauptaspekten zugeordnet und gewichtet: dem Produkt und der Organisation selbst [Al12]. TSCHANDL UND POSCH behandeln die Integration der ökologischen Komponente in das Controlling von Unternehmen [TP12]. Die Autoren XING ET AL. beschreiben Faktoren, welche zur Bewertung von Nachhaltigkeit in PSS herangezogen werden können [XWQ13]. Mit dem Verweis auf HUBBARD, der einen umfassenden Katalog an Treibern für die Nachhaltigkeit präsentiert [Hu06], beschreiben ABRAMOVICI ET AL. einen Ansatz für die gewichtete Beurteilung und Überwachung von Nachhaltigkeit in PSS [Ab14].

Ein weiterer Aspekt der Nachhaltigkeit ist die Tatsache, dass die *einzelnen Dimensionen* nicht unabhängig voneinander betrachtet werden sollten. Einen ersten Ansatz dafür skiz-

zieren AZAPAGIC UND PERDAN [AP00]. Auch WANG UND LIN behandeln die Wechselbeziehungen der einzelnen Dimensionen [WL07]. Weiterführend beschäftigen sich LEE ET AL. mit multidimensionalen und dynamischen Einflüssen von PSS auf die Nachhaltigkeit. Anhand eines öffentlichen Fahrrad-Sharing Systems beschreiben sie die möglichen unterschiedlichen Schnittmengen und Einflüsse der drei Dimensionen der Nachhaltigkeit unter- und aufeinander [LK10].

Zwar konnten im Zuge der Literaturrecherche einige Beispiele für *Reifegradmodelle* identifiziert werden (vgl. 4.1), jedoch wurde keine Arbeit gefunden, die sich mit der Bestimmung des Reifegrades von Nachhaltigkeit in der hybriden Wertschöpfung beschäftigt. Aus diesem Grund wird als Ausgangsbasis für die Entwicklung eines neuen Modells eine systematische Vorgehensweise zur Entwicklung von Reifegradmodellen im IT-Management von KNACKSTEDT ET AL. herangezogen. Durch das erarbeitete Vorgehensmodell soll eine allgemeingültige Methode zur Entwicklung von Reifegradmodellen ermöglicht werden [BKP09].

Abschließend lässt sich anmerken, dass in umfangreichem Maße weitere gefundene Literatur, bspw. zur *Entwicklung und dem Design von nachhaltigen PSS,* darstellt, die aber keinen direkten Implikationen auf die Entwicklung des Modells lieferten. Jedoch können auch aus der hier nicht weiter analysierten Literatur Erkenntnisse vor allem für spezifische Branchen für weitere Forschungsschritte gewonnen werden.

3 Methodisches Vorgehen zur Modellentwicklung

Zur Entwicklung des Reifegradmodells wird eine Vorgehensweise in Anlehnung an KNACKSTEDT ET AL. bzw. BECKER ET AL. angewendet, welche in Abb. 2 dargestellt ist. Diese orientiert sich an den Richtlinien zur Durchführung von Design Science Research, dessen Ziel es ist, neuartige IT-Artefakte in Form von Modellen oder Methoden zu entwickeln, um innovative Möglichkeiten und Fähigkeiten zu schaffen, Probleme besser zu lösen [BKP09, KPB09].

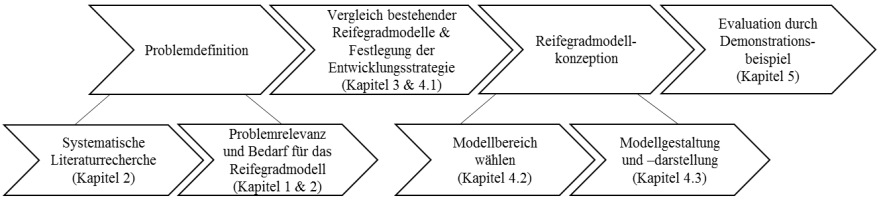


Abb. 2: Vorgehensweise bei der Entwicklung des Reifegradmodells

4 Entwicklung des Reifegradmodells

4.1 Vergleich bestehender Modelle

Wie bereits in der Literaturanalyse festgestellt, gibt es bisher noch keine Ansätze für ein Reifegradmodell, welches die Nachhaltigkeit in hybrider Wertschöpfung misst. Aufgrund dieser Nichtexistenz eines geeigneten Modells, muss ein neues Reifegradmodell entwickelt werden. Dabei kann jedoch auf Bausteine anderer bestehender Modelle zurückgegriffen werden. Deshalb werden im Folgenden in komprimierter Form die Arbeiten von CAGNIN ET AL., KARNI UND KANER und BLECK ET AL. dargestellt und analysiert. Die genannten Arbeiten werden vor allem bezogen auf den *Inhalt des Reifegradmodells*, die *Bewertung der einzelnen Reifegrade* und die *Darstellung der Ergebnisse* verglichen. Diese drei Kategorien werden in den folgenden Kapiteln wieder aufgegriffen.

CAGNIN ET AL. beschreiben in einem ersten Ansatz ein Reifegradmodell für die Entwicklung von Unternehmen hin zu nachhaltigen und innovativen Geschäftsstrategien. Dabei wird der Fokus vor allem auf die Bewertung der verschiedenen Reifegradstufen gelegt. Es werden ausführlich fünf verschiedene Stufen der Reife beschrieben und hierbei u.a. auf die unterschiedlichen Bereiche Strategie, Technologie und Partnerschaften eingegangen [CLB05]. Eine Form der Darstellung der Ergebnisse wird in der Arbeit nicht vorgeschlagen. Das von BLECK ET AL. erarbeitete Reifegradmodell zur Messung von Nachhaltigkeit für das Sustainable Supply Chain Management (SSCM) orientiert sich ebenso wie die vorliegende Arbeit an der Vorgehensweise von BECKER ET AL. Aufgrund dessen beinhaltet die Arbeit u.a. einen übersichtlichen Vergleich bestehender Reifegradmodelle. Des Weiteren werden umfassend die einzelnen Reifegradstufen des SSCM herausgearbeitet und entlang der unterschiedlichen Stufen der Wertschöpfungs- und Lieferkette aufgeschlüsselt. Zudem erfolgt ein anschaulicher Vorschlag für die Darstellung der Reifegrade der einzelnen Stufen der Lieferkette. Diese werden als Dimensionen in einem Radar Chart, auch Netz- oder Sterndiagramm genannt, visualisiert [BWT11]. Anhand eines Reviews analysieren KARNI UND KANER den Einsatz von Reifegradmodellen und deren Anwendung in der hybriden Wertschöpfung, um ein Reifegradmodell für PSS zu entwickeln. Dabei werden u.a. auch ausführlich die verschiedenen möglichen Reifegradstufen und Faktoren der Stufen verglichen. Auf diese soll im folgenden Abschnitt bei der Bewertung der Reifegrade des Einsatzes von PSS-Lösungen zurückgegriffen werden [KK13].

4.2 Definition der Dimensionen

Neben den klassischen drei Dimensionen *Ökonomisch*, *Ökologisch* und *Sozial* sind ebenso die Schnittmengen zu betrachten. Dabei ist u.a. die *ökonomisch-ökologische* Dimension zu nennen, welche Aspekte umfasst, die gleichzeitig eine Verbesserung der finanziellen Situation und den Schutz der Umwelt zur Folge haben. Die *ökonomisch-soziale* Dimension vereint wiederum den finanziellen Nutzen und eine positive gesellschaftliche Ent-

wicklung miteinander. Kriterien, welche unter der *ökologisch-sozialen* Dimension zusammengefasst sind, verbessern sowohl die Auswirkungen auf die Natur, als auch das gesellschaftliche und individuelle Leben.

Die folgenden ausgewählten möglichen Ausprägungen der einzelnen Dimensionen wurden in Anlehnung an [Ab14, AP00, Sc09a, WL07, BWT11, CF11, ER01, Gl13, Hu06, Ko10, Pi10, Sc09b] zusammengetragen (vgl. Tab. 1)

Dimension	*Mögliche Ausprägungen*
Ökonomisch	Umsatzmaximierung, Gewinnmaximierung, Added Value (z.B. EVA oder Sustainable Value Added), Renditesteigerung (z.B. Umsatzrendite, Eigenkapitalrendite), Reaktionsvermögen auf Marktänderungen, Steigerung des Marktanteils, Innovationsmanagement, Forschungsaufwendungen, Zukunftsinvestitionen, Standardisierung, Modularisierung, Instandhaltungssystem, Steuerzahlungen, Produktpreise, Beitrag zum Bruttoinlandsprodukt, Maßnahmen gegen Industriespionage
Ökonomisch-ökologisch	Reduzierung des CO2-Ausstoßes, Einsparung beim Stromverbrauches, Einsparung beim Wasserverbrauches, Glaubhaftes Umweltmarketing, Sicherstellung der Ressourcenversorgung, Kurze Liefer- und Transportwege, Produktverpackungen vermeiden bzw. nachhaltig gestalten, Produktlanglebigkeit, Demontagemöglichkeit von Produkten, Wiederverwendbarkeit und Weiterverkaufsmöglichkeit von Produkten, Erweiterungsfähigkeit von Produkten, Kraftstoffverbrauch der Fahrzeugflotte, Minimierung des Ausschusses in der Produktion
Ökonomisch-sozial	Work-Life-Balance der Mitarbeiter, Gesundheit der Mitarbeiter (Reduzierung der Krankheitstage), Arbeitsunfälle , Beschäftigungserhalt, Umfassendes Angebot von Ausbildungsplätzen, Investition in Humankapital (Weiterbildungsmaßnahmen), Umfassendes Wissensmanagement, Einbindung aller Generationen, Altersteilzeitkonzepte, Arbeitsplatzkonzepte (u.a. Beleuchtung, Belüftung), Glaubhaftes Sozialmarketing, Beziehungsmanagement, Risikokontrolle, Bekämpfung der Korruption, Langfristig angelegte Anreizsysteme für Mitarbeiter, Ausführliche Berichterstattung, Länge der Liefer- und Serviceverträge, Produktsicherheit, Produktnutzen für die Gesellschaft, Einkommensunterschiede im Unternehmen, Unternehmens- und branchenübergreifende Zusammenarbeit, Betriebliche Altersvorsorge, Schutz von Kundendaten
Ökologisch	Berücksichtigung erneuerbarer Energien, Umsetzung von Green-IT-Lösungen, Einsatz von Umweltmanagementsystemen, Umweltzertifizierungen, Sachgerechte Entsorgung von Elektronikmüll, Nutzung von Energiesparlampen, Verwendung recycelter Materialien, Zusammenarbeit mit Umweltverbänden, Beitrag zum Klimaschutz bzw. Kampf gegen die Erderwärmung, Maßnahmen gegen Versäuerung, Eigene Stromproduktion (durch Photovoltaik, Windkrafträder oder Bio-Gasanlagen)
Ökologisch-sozial	Mitarbeiterkompetenz im Bereich Klimaschutz, Mitarbeiterkompetenz im Bereich Müllentsorgung und Recycling, Förderung von Fahrgemeinschaften, Unterstützung von Sharing-Konzepten, Minimierung der Verwendung giftiger Substanzen, Überwachung der (ökologischen und sozialen) Standards der Zulieferer und Abnehmer, Produkte und Dienstleitungen fördern nachhaltigen Konsum, Einsatz für den Klimaschutz in der lokalen Gesellschaft
Sozial	Steigerung der Mitarbeiterzufriedenheit, Förderung des Gemeinschaftsgedankens, Faire Arbeitsverträge, Einkommensverteilung und -unterschiede, Angebote für Familien, Unternehmensinterne Kinderbetreuung, Verantwortliches Handeln im Bereich Datenschutz und -sicherheit, Spenden für wohltätige Zwecke, Achtung kultureller Vielfalt (der Mitarbeiter), Einsatz für die lokale Gesellschaft, Einsatz gegen Kinderarbeit (auch in der Zuliefererkette), Mitsprache und Entscheidungsrecht von Mitarbeitern, Respekt der Privatsphäre, Frauen in Führungspositionen, Maßnahmen zur Gleichberechtigung der Geschlechter, Maßnahmen gegen Diskriminierung, Barrierefreies Arbeiten, Chancengleichkeit, Flexible Arbeitszeiten (z.B. Gleitzeit)

Tab. 1: Ausprägungen der Dimensionen der Nachhaltigkeit

Bei den Ausprägungen wird deutlich, dass vor allem im Bereich Ökonomisch, Ökonomisch-ökologisch sowie Ökologisch, Ausprägungen in der Literatur genannt werden, die erst durch das Angebot der hybriden Leistung möglich sind. Beispiele hierfür ist die Demontagemöglichkeit von Produkten, Wiederverwendbarkeit und Weiterverkaufsmöglichkeit von Produkten. Des Weiteren gibt es Ausprägungen deren Wert durch das hybride

Angebot, wie die Reduzierung des CO2-Ausstoßes durch ständige und korrekte Wartung gesteigert werden kann.

4.3 Ausprägungen der Reifegrade

Zur Beurteilung des Reifegrades der *(1) Nachhaltigkeit* müssen die Ausprägungen der sechs Dimensionen einzeln ausgewertet werden. Als Einordnungshilfe, auf deren Basis später auch die Beurteilung von Experten oder die Selbstevaluation anhand eines Fragebogens erfolgen kann, dient die folgende Tab. 2.

Stufe	*Beschreibung und Merkmale*	
	(1) Nachhaltigkeit	*(2) PSS*
0 - inaktiv	Das Prinzip der Nachhaltigkeit findet keinerlei Beachtung und die ökologische bzw. soziale Dimension ist nicht bekannt. Die Verantwortlichen zeigen kein Interesse an einer nachhaltigen Entwicklung und Geschäftspolitik entlang dieser Dimension bzw. Ausprägung. Gesetzliche Vorschriften werden nicht umgesetzt.	Es bestehen keine Kenntnisse über das Prinzip der hybriden Wertschöpfung. Deshalb werden keine PSS eingesetzt, um die Nachhaltigkeit zu verbessern.
1 - ad hoc	Es existieren Kenntnisse von der Nachhaltigkeit als solche, erste nachhaltige Ideen und Vorschläge. Diese Kenntnisse werden jedoch nur vereinzelt deutlich und die Ausprägungen werden nur begrenzt wahrgenommen. Nachhaltigkeitsaktivitäten sind vereinzelt erkennbar.	Das Prinzip der hybriden Wertschöpfung ist bekannt. Prozesse werden isoliert betrachtet. PSS bestehen im Unternehmen nur zufällig und ungeplant.
2 – regel-konform und bewusst	Gesetzliche Vorschriften werden erfüllt. Es werden erste nachhaltige Ziele gesetzt und die Ausprägungen werden als Nachhaltigkeitskriterien bewusst wahrgenommen. Dabei werden aber vor allem kurzfristige, nicht auf Dauer angelegte Effizienzsteigerungen der nachhaltigen Dimension bzw. Ausprägung verfolgt. Mögliche Hürden auf dem Weg zu mehr Nachhaltigkeit werden identifiziert.	Der Einsatz von PSS wird verstanden und bewusst in Teilen des Unternehmens eingesetzt, um die Nachhaltigkeit zu verbessern. Der Einsatz ist jedoch nur kurzfristig geplant und erfasst nicht alle Phasen des Produktlebenszyklus. Der Einsatz von IT erfolgt nicht über das notwendige Maß hinaus.
3 - aktiv und etabliert	Nachhaltigkeitsaktivitäten werden isoliert geplant und dokumentiert. Gesetzliche Vorschriften werden selbstgewählt übertroffen. Vorgehen im Bereich Nachhaltigkeit wird nach Innen und Außen kommuniziert und aktiv verfolgt. Die Notwendigkeit zur Verbesserung der nachhaltigen Ausprägung hat sich etabliert.	PSS-Lösungen haben sich unternehmensweit etabliert, werden jedoch isoliert geplant und dokumentiert. Es existiert ein umfassender Einsatz von IT zur Überwachung der Prozesse. Die Bedürfnisse der Konsumenten werden erkannt. Es wird eine aktive Steigerung der Nachhaltigkeit der betrachteten Ausprägung angestrebt.
4 - integriert & quantitativ gesteuert	Es erfolgt eine simultane Planung der Nachhaltigkeitsaktivitäten entlang der Ausprägungen. Das Streben nach höherer Nachhaltigkeit ist in die Unternehmensstrategie integriert. Ökologisches bzw. soziales Engagement ist langfristig ausgerichtet und wird quantitativ überprüft und gesteuert.	Der Einsatz der hybriden Wertschöpfung gehört zur festgesetzten Unternehmensstrategie. PSS werden bereichsübergreifend gesteuert und sind in allen Phasen des Produktlebenszyklus zu finden. Die IT wird prozessintegriert eingesetzt und die PSS sind quantitativ gesteuert.
5 - proaktiv und optimierend	Das nachhaltige Handeln ist intrinsisch motiviert und gehört zu einem festen Bestandteil der Unternehmenskultur. Es werden Standards in den einzelnen Ausprägungen der Nachhaltigkeit gesetzt und es wird eine Vorreiterrolle eingenommen. Die Nachhaltigkeit wird kontinuierlich anhand der erhobenen quantitativen Daten verbessert.	Konsumentenbedürfnisse werden transformiert und neue innovative PSS entwickelt. Die langfristige Entwicklung des Unternehmens ist an die angebotenen und genutzten, nachhaltigen PSS gebunden. Es wird eine Führungsposition im Bereich der hybriden Wertschöpfung eingenommen.

Tab. 2: Beschreibung und Merkmale der Reifegradstufen

Die Auswertung erfolgt in *zwei Schritten*: Zunächst wird die Nachhaltigkeit in den sechs

Dimensionen bestimmt und anschließend beurteilt, in welchem Maß PSS eingesetzt werden, um diesen bestimmten Reifegrad zu erreichen. Bei Simulation von PSS-Lösungen kann später zunächst Schritt 2 ausgeführt werden. Die folgenden Merkmale wurden in Anlehnung an [BWT11, CLB05, KK13, Za10] erarbeitet. Zur Beurteilung des Reifegrades der Nachhaltigkeit und der hybriden Wertschöpfung wurden sechs aufeinander aufbauende Reifegradstufen gewählt, welche fließend bzw. stetig ineinander übergehen. Aus diesem Grund können bei der Aggregation in den sechs übergeordneten Dimensionen der Nachhaltigkeit auch Reifegrade erreicht werden, die zwischen zwei der genannten Stufen liegen.

4.4 Darstellung der Reifegrade

Zur vergleichenden Darstellung der Reifegrade der Nachhaltigkeit und des Einsatzes von PSS wird die Nutzung eines Netz- bzw. Sterndiagramms vorgeschlagen. Der Vorteil ist, dass Netzdiagramme relativ unkompliziert, bspw. in Microsoft Excel, erstellt werden können, also keine spezielle Software zur Darstellung angeschafft werden muss. Des Weiteren stellen Netzdiagramme ein passendes Werkzeug zur Visualisierung von Reifegraden dar, weil sich die verschiedenen Dimensionen der Nachhaltigkeit auf einen Blick beurteilen lassen und der Einsatz von PSS deutlich wird [BWT11].

Abb. 3 zeigt die beispielhafte Darstellung eines in Excel erstellten Netzdiagrammes mit den sechs Dimensionen der Nachhaltigkeit und den beiden Aspekten Nachhaltigkeit und PSS. Für diese beiden Aspekte wurden die Reifegrade innerhalb der einzelnen Dimensionen anhand der Ausprägungen des Kataloges bestimmt.

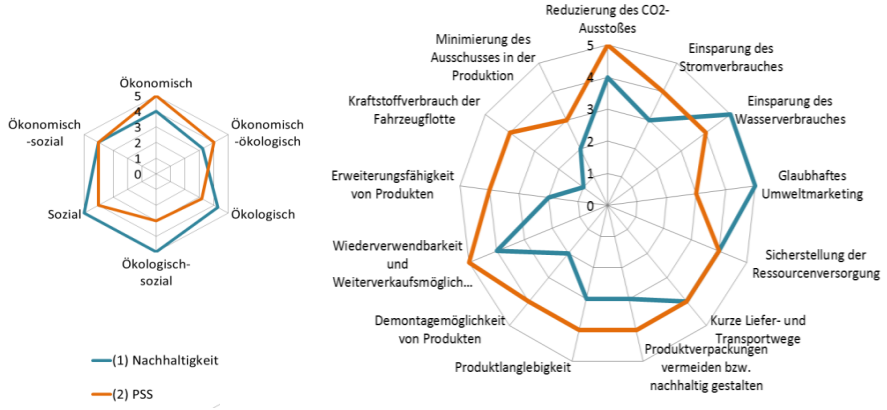


Abb. 3: Netzdiagramm zur Darstellung der Reifegrade und Subnetzdiagramm zur Dimension Ökonomisch-ökologisch

Des Weiteren können zur spezifischeren Auswertung und Darstellung Subnetzdiagramme erstellt werden, welche die Beurteilung der einzelnen Ausprägungen der Nachhaltigkeit

visualisieren. Hierfür lässt sich das Netzdiagramm um beliebig viele weitere Dimensionen erweitern.

Zudem kann bei Standardisierung des Modells der Vergleich verschiedener Unternehmen oder Unternehmensdivisionen vorgenommen werden, deren Ergebnisse dann vergleichend in einem Netzdiagramm abgebildet und interpretiert werden können. Des Weiteren können auch vergleichende Analysen verschiedener Zeitpunkte durchgeführt werden, um herauszuarbeiten, wie sich das Unternehmen bzw. Produkte und Dienstleistungen mit der Zeit entwickelt haben.

5 Evaluation durch Demonstrationsbeispiel

Das entwickelte Modell kann als Grundlage für eine umfassende Studie zur Messung von Nachhaltigkeit in Unternehmen genutzt werden. In dieser Messung soll identifiziert werden, wo bereits PSS-Lösungen eingesetzt werden und welche Auswirkungen diese auf die Nachhaltigkeit von Unternehmen haben. Dazu ist zunächst zu evaluieren, ob das Modell grundsätzlich anwendbar ist und welche Implikationen sich auf die prototypische Entwicklung ergeben. Hierzu wird als Evaluationsmethode das Demonstrationsbeispiel [RSB09] gewählt.

Zur Bestimmung der einzelnen Reifegrade wurde ein Fragebogen, in Anlehnung an den Fragenkatalog nach BLECK ET AL. für sein SSCM-Modell, konzipiert. Ein Ausschnitt ist beispielhaft in Tab. 3 dargestellt. Das Ziel ist es, mit unterschiedlichen Fragen eine möglichst genaue Beurteilung des Reifegrades zu erhalten. Aus den abgegebenen Antworten und den zugeordneten Reifegradstufen [0], [1], [2], [3], [4] oder [5] kann unter Einbezug der Gewichtung dann ein Reifegrad berechnet werden.

Als beispielhafte Ausprägung wird aus dem Bereich Ökonomisch-ökologisch *Wiederverwendbarkeit und Weiterverkaufsmöglichkeit von Produkten* gewählt. Anhand des Fragebogens stuft das Beispielunternehmen, welches Großküchen verkauft, seine Nachhaltigkeit wie folgt ein: 1*0,4 + 2*0,15 + 3*0,15 + 2*0,15 + 2*0,15 = 1,75. Als Begründung dafür wird angegeben, dass das Unternehmen defekte Produkte zurücknimmt und recycelt, wenn diese eingeschickt werden, dies aber nur ad-hoc im Fall eines Eintreffens geschieht. In Bezug auf die PSS-Lösung wird folgende Einschätzung getroffen: 1*0,4 + 3*0,1 + 3*0,15 + 2*0,15 + 2*0,1 = 1,65. Die Gründe hierfür sind, dass das Unternehmen für den Kunden über die Internetseite zwar anbietet, Produkte zurückzunehmen, dies aber bisher nur für defekte Produkte gilt. Die Einschätzung des Reifegrads der hybriden Wertschöpfung und der Nachhaltigkeit dieser Strategie stimmen weitestgehend überein. Wenn sich das Unternehmen entschließen würde, Ihr Produkt mit Sensoren und GPS auszustatten (Frage 2d: Umfassender Einsatz), könnten sie über den gesamten Produktlebenszyklus (Frage 2c: ... komplett erfasst) erfahren, wo sich ihr Produkt befindet, und welchen Status dies hat. Automatisch könnte regelmäßig der Kunde angeschrieben werden, ob er das Produkt auch im funktionieren Zustand noch benötigt, oder ob dies durch neuere Geräte ersetzt und das alte zurückgenommen werden soll.

Fragebogenabschnitt: Auswertung einer möglichen Ausprägung	*Gewichtung und Reifestufe*	*Beispiel*
(1) Wie schätzen Sie den Reifegrad der Ausprägung in Bezug auf die Nachhaltigkeit ein?	40%	
Inaktiv	[0]	
Ad hoc	[1]	X
Regelkonform und bewusst	[2]	
Aktiv und etabliert	[3]	
Integriert und quantitativ gesteuert	[4]	
Proaktiv und optimierend	[5]	
(1a) Inwiefern werden gesetzliche Vorschriften umgesetzt?	15%	
... werden nicht immer eingehalten.	[0,5]	
... werden eingehalten.	[2]	X
... werden bewusst übertroffen.	[3]	
... es werden neue wegweisende Standards ausgearbeitet.	[5]	
(1b) Wie erfolgt die Planung der Nachhaltigkeit in der Ausprägung?	15%	
Keine Planung	[1]	
Isolierte Planung	[3]	X
Simultane Planung	[4]	
(1c) Was ist der angestrebte Zeithorizont der Ausprägung?	15%	
Kurzfristige Effizienzsteigerung	[2]	X
Mittelfristig	[3]	
Langfristig	[4,5]	
(1d) Die Notwendigkeit der nachhaltigen Verbesserung der Ausprägung ist in der Unternehmenskultur...	15%	
... nicht verankert.	[0]	
... bewusst.	[2]	X
... quantitativ gesteuert.	[4]	
... intrinsisch verankert.	[5]	
(2) Wie schätzen Sie den Reifegrad der Ausprägung in Bezug auf den Einsatz von PSS ein?	40%	
Inaktiv	[0]	
Ad hoc	[1]	X
Regelkonform und bewusst	[2]	
Aktiv und etabliert	[3]	
Integriert und quantitativ gesteuert	[4]	
Proaktiv und optimierend	[5]	
(2a) Wie erfolgt die Planung des PSS-Einsatzes in der Ausprägung?	10%	
Keine Planung	[1]	
Isolierte Planung	[3]	X
Simultane Planung	[4]	
(2b) Was ist der angestrebte Zeithorizont beim Einsatz der PSS-Lösung in der Ausprägung?	10%	
Kurzfristige Effizienzsteigerung	[2]	
Mittelfristig	[3]	X
Langfristig	[4,5]	
(2c) Die Phasen des Produktlebenszyklus werden...	15%	
... nicht betrachtet.	[0]	
... nicht alle einbezogen.	[2]	X
... komplett erfasst.	[4]	
... optimiert.	[5]	
(2d) In welchem Ausmaß erfolgt der Einsatz von IT in der Ausprägung?	15%	
Keine IT	[0]	
Nur im notwendigen Maße	[2]	X
Umfassender Einsatz	[4,5]	
(2e) Die Notwendigkeit des nachhaltigen Einsatzes von PSS in der Ausprägung ist in der Unternehmenskultur...	10%	
... nicht verankert.	[0]	
... bewusst.	[2]	X
... quantitativ gesteuert.	[4]	
... intrinsisch verankert.	[5]	

Tab. 3: Fragebogen zur Bestimmung des Reifegrades

Eine Implementierung dieser Handlungsempfehlung würde die Nachhaltigkeit stärken. Durch die strategische Planung des PSS-Einsatzes würden auch die Antworten auf die Fragen der Nachhaltigkeit (Frage 1: integriert, Frage 1a: neue wegweisende Standards, etc.) einen höheren Reifegrad ergeben. Somit kann das Unternehmen die Auswirkung der PSS-Maßnahmen auf die Nachhaltigkeit direkt durch den Bericht ablesen. Das Demonstrationsbeispiel zeigt, dass eine Umsetzung grundsätzlich möglich ist und birgt folgende Implikationen für die prototypische Implementierung: (1) Die Abhängigkeiten der Fragen zur PSS-Lösung auf die Nachhaltigkeit müssen hinterlegt werden, sodass die Antworten zur Nachhaltigkeit automatisch angepasst werden (vgl. Beispielhafte Handlungsempfehlung) (2) Die Gewichtung der einzelnen Fragen muss individuell für den spezifischen Unternehmenskontext anpassbar sein, da ggf. gesetzliche Vorschriften in einigen Unternehmen wichtiger sind, als bei dem Demonstrationsbeispiel, (3) es sollte eine Simulation möglicher Maßnahmen bzw. ein Vergleich dieser möglich sein.

6 Fazit

Die in dieser Arbeit durchgeführte systematische Literaturrecherche und -analyse hat gezeigt, dass zwar beispielhafte Ansätze zur Messung von Nachhaltigkeit in PSS existieren, bisher jedoch kein umfassendes Modell zur Messung und zum Bericht des Reifegrads von Nachhaltigkeit in hybrider Wertschöpfung entwickelt wurde. Aufgrund dessen wurde ein neues Modell konzipiert, welches als Grundlage für weiterführende, empirische Forschung genutzt werden kann. Hierbei ist zu untersuchen, ob und wie es zu Abweichungen der Reifegrade (hybrid > nachhaltig, nachhaltig > hybrid) kommt, um im nächsten Schritt die automatischen Auswirkungen im Tool implementieren zu können.

Dabei hat sich vor allem gezeigt, dass es wichtig ist, die Schnittmengen der Dimensionen der Nachhaltigkeit in die Analyse mit einzubeziehen. Es konnte ein Katalog an Ausprägungen in den sechs Dimensionen der Nachhaltigkeit erstellt werden. Darüber hinaus wurden Merkmale und Beschreibungen für die einzelnen Stufen der Reife in Nachhaltigkeit und hybrider Wertschöpfung identifiziert. Als intuitive Darstellungsmethode wurde die Verwendung eines Netzdiagramms gewählt. Die parallele Auswertung von Nachhaltigkeit und PSS offeriert Unternehmen Potential zur Anpassung ihrer Strategie. Wie bereits mehrfach erwähnt, muss im Anschluss an diese Arbeit eine ausführliche empirische Validierung des Reifegradmodells erfolgen.

Es hat sich gezeigt, dass der Einsatz von PSS-Lösungen in den meisten Fällen zu höherer Nachhaltigkeit führt. Dennoch kann die unbedachten Einführung von PSS-Lösungen auch zu einer geringeren Nachhaltigkeit einzelner Dimensionen führen, wie z.B. beim Leasing [Tu04]. Die Unternehmen müssen das Konzept der hybriden Wertschöpfung und die Auswirkungen von PSS auf die Nachhaltigkeit verstehen. In dieser Arbeit wurde in Form des entwickelten Reifegradmodells eine Grundlage für ein Werkzeug gelegt, mit dem man diesen Zusammenhang analysieren kann.

Zur Implementierung des vorgeschlagenen Modells müssen im nächsten Schritt die Konstruktionsmerkmale sowohl für die technische Implementierung als auch die organisatorische Verankerung im Unternehmen bestimmt werden. Das vorgestellte Modell liefert, neben den anhand des Demonstrationsbeispiels hergeleiteten Anforderungen, weitere Implikationen für die Implementierung. Es ist zu beachten, dass die Stakeholder (u.a. Abteilungen wie Dienstleistungsentwicklung, Produktentwicklung, Prozessmanagement, Controlling, Strategische Entscheider/ Unternehmensführung) an unterschiedlichen Positionen im Unternehmen verteilt sind. Es muss eine Lösung geschaffen werden, die zentral zugänglich ist, bspw. durch einen webbasierten Zugang. Der Input der Daten kann sowohl in Selbstevaluation [BKP09], als auch automatisiert aus bestehenden Anwendungssystemen, bspw. dem ERP-System, geschehen. Hierzu sind sowohl eine Eingabemaske (u.a. zur Beantwortung der in Tab. 3 dargestellten Fragen) als auch eine Schnittstelle für die Anwendungssysteme zu schaffen.

Schlussfolgernd ist festzuhalten, dass Unternehmen größere Wettbewerbschancen in der Zukunft haben werden, wenn sie sich möglichst früh intensiv mit der hybriden Wertschöpfung auseinandersetzen. Deshalb stellt das in diesem Beitrag präsentierte Reifegradmodell eine Möglichkeit dar, um die eigene Positionierung in Bezug auf die Nachhaltigkeit in der hybriden Wertschöpfung zu messen und zu berichten. Darauf aufbauend kann vor der Implementierung einer hybriden Maßnahme die direkte Auswirkung auf die sechs Dimensionen der Nachhaltigkeit abgelesen werden.

Literaturverzeichnis

[Ab14] Abramovici, M. et.al.: PSS Sustainability Assessment and Monitoring Framework (PSS-SAM) – Case Study of a Multi-module PSS Solution. In: Procedia CIRP Bd. 16 (2014), S. 140–145

[Al12] Allen Hu, H. et.al.: Development of sustainability evaluation model for implementing product service systems. In: International Journal of Environmental Science and Technology Bd. 9 (2012), Nr. 2, S. 343–354

[AP00] Azapagic, A; Perdan, S: Indicators of sustainable development for industry: a general framework. In: Trans IChemE Bd. 78 (2000), Nr. 4, S. 243–261

[BKP09] Becker, J.; Knackstedt, R.; Pöppelbuß, J.: Entwicklung von Reifegradmodellen für das IT-Management. In: Wirtschaftsinformatik Bd. 51 (2009), Nr. 3, S. 249–260

[Ba07] Baines, T. S. et.al.; u. a.: State-of-the-art in product-service systems. In: Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture Bd. 221 (2007), Nr. 10, S. 1543–1552

[BWT11] Bleck, F.; Wittstruck, D.; Teuteberg, F.: Entwicklung und Validierung eines Reifegradmodells für das Sustainable Supply Chain Management. In: INFORMATIK 2011 - Informatik schafft Communities: 41. Jahrestagung der Gesellschaft für Informatik. Berlin, 2011, S. 1–14

[BT13] Boehm, M.; Thomas, O.: Looking beyond the rim of one's teacup: a multidisciplinary literature review of Product-Service Systems in Information Systems, Business Management, and Engineering & Design. In: Journal of Cleaner Production Bd. 51, Elsevier Ltd

(2013), S. 245–260

[Br09] von Brocke, J. et.al.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: 17th European Conference on Information Systems (2009), S. 2206–2217

[Bu09] Bundesministerium für Umwelt Naturschutz und Reaktorsicherheit: Nachhaltigkeits-berichterstattung: Empfehlungen für eine gute Unternehmenspraxis. Berlin, 2009

[CLB05] Cagnin, C. H.; Loveridge, D.; Butler, J.: Business Sustainability Maturity Model. In: Business Strategy and the Environment Conference. Leeds, 2005, S. 1–15

[CL10] Colen, P.; Lambrecht, M.: Product service systems as a vehicle for sustainability : Exploring service operations strategies. In: 3rd International Conference on Information Systems, Logistics and Supply Chain Creating value through green supply chains. Casablanca, 2010, S. 1–10

[CL09] Coley, F. J. S.; Lemon, M.: Exploring the design and perceived benefit of sustainable solutions: a review. In: Journal of Engineering Design Bd. 20 (2009), Nr. 6, S. 543–554

[CF11] Cunha, C., Leonardo, A.; Fensterseifer, J. E.: Corporate Sustainability Measure From An Integrated Perspective: The Corporate Sustainability Grid. In: International Journal of Business Insights & Transformation Bd. 3 (2011), Nr. 3, S. 44–53

[ER01] Epstein, M. J.; Roy, M.-J.: Sustainability in Action: Identifying and Measuring the Key Performance Drivers. In: Long Range Planning Bd. 34 (2001), Nr. 5, S. 585–604

[Fi09] Finnveden, G. et.al.: Recent Developments in Life Cycle Assessment. In: Journal of Environmental Management Bd. 91 (2009), Nr. 1, S. 1–21

[Go99] Goedkoop, M.J et.al.: Product Service systems, Ecological and Economic Basics. The Hague: VROM/EZ, 1999

[Gl13] Global Reporting Initiative: G4 Leitlinien zur Nachhaltigkeitsberichterstattung. Amsterdam, 2013

[HGR09] Hansen, E. G.; Grosse-Dunker, F.; Reichwald, Ralf: Sustainability Innovation Cube — a Framework To Evaluate Sustainability-Oriented Innovations. In: International Journal of Innovation Management Bd. 13 (2009), Nr. 04, S. 683–713

[Hu06] Hubbard, G.: Sustainable organisation performance : Towards a practical measurement system Theory of the organisation determines performance measurement approach Recent developments of organisational performance measuring systems. In: Monash Business Review Bd. 2 (2006), Nr. 3, S. 1–19

[KK13] Karni, R.; Kaner, M.: A Review of Maturity Models and their Application to PSS: Towards a PSS Maturity Model. In: Shimomura, Y.; Kimita, K. (Hrsg.): The Philosopher's Stone for Sustainability. Berlin Heidelberg: Springer, 2013, S. 393–398

[KPB09] Knackstedt, R.; Pöppelbuß, J.; Becker, J.: Vorgehensweise zur Entwicklung von Reifegradmodellen. In: Hansen, H. R.; Karagiannis, D.; Fill, H.-G. (Hrsg.): 9. Internationale Tagung Wirtschaftsinformatik: Business Services: Konzepte, Technologien, Anwendungen. Wien, 2009, S. 535–544

[KPW08] Knackstedt, R.; Pöppelbuß, J.; Winkelmann, A.: Integration von Sach- und Dienstleistungen – Ausgewählte Internetquellen zur hybriden Wertschöpfung. In: Wirtschaftsinformatik Bd. 50 (2008), Nr. 3, S. 235–247

[Ko10] Koltun, P.: Materials and sustainable development. In: Progress in Natural Science: Materials International Bd. 20, Chinese Materials Research Society (2010), S. 16–29

[LK10] Lee, S W; Kim, Y S: A Product-Service Systems Design Method Integrating Service Function and Service Activity and Case Studies. In: Proceedings of the 2nd CIRP International Conference on Industrial Product Service Systems, 2010, S. 275–282

[Mo02] Mont, O.: Clarifying the concept of product-service system. In: Journal of Cleaner Production Bd. 10 (2002), Nr. 3, S. 237–245

[My14] Mylan, J.: Understanding the diffusion of Sustainable Product-Service Systems: Insights from the sociology of consumption and practice theory. In: Journal of Cleaner Production, Elsevier Ltd (2014), S. 1–8

[NTF14] Nüttgens, M.; Thomas, O.; Fellmann, M. (Hrsg.): Dienstleistungsproduktivität: Mit mobilen Assistenzsystemen zum Unternehmenserfolg. Wiesbaden: Springer Fachmedien Wiesbaden, 2014

[Pi10] Pigosso, D.C.A et.al.: Is the Industrial Product-Service System really sustainable? In: Sakao, T.; Larsson, T.; Lindahl, M. (Hrsg.): 2nd CIRP IPS Conference 2010. Linköping, 2010, S. 59–65

[RSB09] Riege, C.; Saat, J.; Bucher, T.: Systematisierung von Evaluationsmethoden in der gestaltungsorientierten Wirtschaftsinformatik. In: Becker, J.; Krcmar, H.; Niehaves, B. (Hrsg.): Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik: Physica-Verlag HD, 2009, S. 69–86

[Ro00] Roy, R.: Sustainable product-service systems. In: Futures Bd. 32 (2000), S. 289–299

[Sc14] Schaltegger, S.: Nachhaltigkeitsberichterstattung zwischen Transparenzanspruch und Management der Nachhaltigkeitsleistung. In: Fifka, M. S. (Hrsg.): CSR und Reporting SE - 2, Management-Reihe Corporate Social Responsibility: Springer Berlin Heidelberg, 2014, S. 21–34

[Sc09a] Schmidt, N.-H. et.al.: Towards a Procedural Model for Sustainable Information Systems Management. In: HICSS 2009, S. 1–10

[Sc09b] Schmidt, N.-H. et.al.: Nachhaltiges Informationsmanagement. In: Wirtschaftsinformatik Bd. 51 (2009), Nr. 5, S. 463–466

[TLN10] Thomas, O.; Loos, P.; Nüttgens, M. (Hrsg.): Hybride Wertschöpfung - Mobile Anwendungssysteme für effiziente Dienstleistungsproduktivität im technischen Kundendienst. Berlin, Heidelberg: Springer, 2010

[TP12] Tschandl, M.; Posch, A. (Hrsg.): Integriertes Umweltcontrolling. 2. Aufl. Wiesbaden: Gabler, 2012

[Tu04] Tukker, A.: Eight Types of Product-Service System: Eight ways to Sustainability? Experiences from Suspronet. In: Business Strategy and the Environment Bd. 13 (2004), S. 246–260

[WL07] Wang, L.; Lin, L.: A methodological framework for the triple bottom line accounting and management of industry enterprises. In: International Journal of Production Research Bd. 45, Taylor & Francis Group (2007), Nr. 5, S. 1063–1088

[XNL13] Xing, Ke; Ness, D.; Lin, F.: A service innovation model for synergistic community transformation: integrated application of systems theory and product-service systems. In: Journal of Cleaner Production Bd. 43, Elsevier Ltd (2013), S. 93–102

[XWQ13] Xing, K.; Wang, H.-F.; Qian, W.: A sustainability-oriented multi-dimensional value assessment model for product-service development. In: International Journal of Production Research Bd. 51 (2013), Nr. 19, S. 5908–5933

[Za10] Zarnekow, Rüdiger et.al.: Planungs- und Steuerungskonzepte eines nachhaltigen Informationsmanagements, 2010

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Nutzung individuell bestimmbarer KPIs im Incident Management zur nachhaltig effektiven Schwachstellen- und Stärkenidentifizierung von IT-Service Landschaften

Volkan Gizli[1], Jorge Marx Gómez[2] und Simon Stephane Ntomthe Tiako[3]

Abstract: In einer durch die Globalisierung sowie durch zahlreiche Unternehmensübernahmen geprägten Informationsinfrastruktur von Großkonzernen steigt zunehmend die Komplexität der IT-Service Landschaften und somit die Übersicht der damit verbundenen Applikationen. Mit diesem Hintergrund wird das Ziel verfolgt, eine Transparenz im Bereich des Application Service Managements (ASM) zu schaffen. Dieses ist z. B. durch die Entwicklung eines Dashboardsystems mittels Implementierung von Kennzahlen aus ITIL zu realisieren. Dadurch lassen sich IT-Services kontrollieren, um kritische Situationen einfacher und schneller zu erkennen, die dementsprechend nachhaltig gesteuert werden können. Mit diesem Problem beschäftigte sich die Forschungskooperation *Dynamic IT Supplier Management* mit der Carl von Ossietzky Universität Oldenburg und einem Unternehmen aus der Automotivbranche, in dem sie ein Dashboardsystem entwickelte, was in diesem Beitrag aufgezeigt werden soll.

Keywords: IT-Service, Lieferantenmanagement, Dashboard, KPI, Nachhaltige Transparenz

1 Nachhaltige Transparenz in IT-Service Landschaften

Die Transparenz in IT-Service Landschaften und die Übersicht der mit den damit verbundenen Applikationen ist von großer Bedeutung, wenn es insbesondere darum geht, kritische Störfälle aufzuspüren und entsprechend zu steuern. Durch die Globalisierung von Großkonzernen steigen stetig die Komplexität der Informationsstrukturen und somit auch die Kosten in der IT. Es wird das Ziel verfolgt, eine bessere Übersicht im Application Service Management (ASM) zu schaffen.
Ein Ansatz für die Kontrolle [FBS12] der Störfälle in den IT-Services bietet die Implementierung von Kennzahlen aus ITIL, die z. B. mittels der Entwicklung eines Dashboardsystems verwendet werden können. „Die Information Technology Infrastructure Libraray (ITIL) ist ein öffentliches Rahmenwerk von „Best Practices"-Lösungen, das weltweit als De-facto-Standard im Bereich des IT-Service Managements gilt.

[1] Department für Informatik, Very Large Business Application, Carl von Ossietzky Universität Oldenburg, 26129 Oldenburg, volkan.gizli@uni-oldenburg.de
[2] Department für Informatik, Very Large Business Application, Carl von Ossietzky Universität Oldenburg, 26129 Oldenburg, jorge.marx.gomez@uni-oldenburg.de
[3] Department für Informatik, Very Large Business Application, Carl von Ossietzky Universität Oldenburg, 26129 Oldenburg, simon.stephane.ntomthe.tiako@uni-oldenburg.de

In der Praxis sind viele Unternehmen an einer klaren Definition von IT-Services gescheitert, was nicht zuletzt der Tatsache geschuldet ist, dass die Servicenehmer IT-Services als eine Gesamtheit wahrnehmen. Sehr häufig besteht aber ein IT-Service aus dem Zusammenspiel mehrerer IT-Funktionen. Diese IT-Funktionen wiederum bestehen aus Hardware-, Software-, Kommunikations- und Dienstleistungskomponenten, werden aber von den Kunden, die diese Services in Anspruch nehmen als eine geschlossene Einheit wahrgenommen."[4] Mit einem solchen Dashboardsystems können kritische Situationen erkannt werden, um sie frühzeitig zu beheben und Konsequenzen nachhaltig zu vermeiden, was somit als Treiber von Nachhaltigkeit im Unternehmenskontext fungieren kann.[5] Das Problem der Intransparenz in den IT-Services und den damit zusammenhängenden Applikationen ist die zunehmende Komplexität, mit der die Übersicht kritischer Störfälle schwer aufzuspüren ist und zu stetig steigenden Kosten führt.[6] Diese zunehmende Komplexität erschwert durch die Unübersichtlichkeit immer mehr die Steuerung der IT-Services. Um diesem Problem entgegenzuwirken, ist es notwendig, kritische Störfälle aus einer solchen IT-Service Landschaft zeitnahe zu erkennen und zu priorisieren, damit diese entsprechend nachhaltig gesteuert werden können. Hierbei wird das Application Service Management (ASM)[7], mit den damit verbundenen Applikationen der IT-Services betrachtet. An dieser Stelle wird der ökonomische Aspekt des Drei-Säulen-Modells der Nachhaltigkeit adressiert.[8] Die Forschungskooperation *Dynamic IT Supplier Management* mit der Carl von Ossietzky Universität Oldenburg und einem Unternehmen aus der Automotivbranche beschäftigte sich mit diesem Problem. Initial wurde dabei folgende Forschungsfrage aufgestellt, mit dessen Problemlösungen sie sich unter anderem beschäftigt: „Wie können Lieferanten und deren Services im IT-Sektor *bewertet, ausgewählt, verglichen* sowie *gesteuert* werden?" Neben den Problemen und den damit verbundenen Methoden zur Bewertung, Auswahl und dem Vergleich, wurde bisher eine Steuerungsmöglichkeit betrachtet, mit dem eine Übersichtlichkeit mittels eines Dashboardsystems, auf denen essentielle Informationen wie Kennzahlen aus dem Incident Management visuell dargestellt werden können. Somit soll folglich eine bessere Kontrolle über IT-Services ermöglicht werden, indem kritische Situationen schnell zu erkennen sind und infolgedessen so früh wie möglich gesteuert werden können, um Zeit einzusparen. Eine dieser Steuerungsmöglichkeiten bietet das Incident Management, mit dem das Ziel bestrebt wird, den eigentlichen IT-Service so schnell wie möglich in irgendeiner Art und Weise wiederherzustellen und somit die negativen Auswirkungen auf das Geschäft so weit wie möglich vorzubeugen. Insgesamt geht es also darum, dass der Verbraucher bzw. Nutzer uneingeschränkt arbeiten kann und der IT-Service seinen vertraglichen Verpflichtungen nachkommen kann.[9] Ein Incident Management befasst sich mit der Erkennung und Aufzeichnung, Klassifizierung und Sofort- und Ersthilfe (unmittelbare Behebung eines Incidents), Untersuchung und Diagnose, Lösung und Wiederherstellung, Abschluss des Incidents, Eigentümer Incident, Überwachung, Verfolgung und Kommunkation.[10] *"Durch das Incident Management wer-*

[4] [GJO09, S. 110]
[5] Vgl. [ZK13, S. 14]
[6] Vgl. [RK09, S. 1]
[7] Vgl. [Wi12, S. 132-133]
[8] [Kr11, S. 63ff]
[9] Vgl. [WS14, S. 26]
[10] Vgl. [WS14, S. 27]

den Störfälle frühzeitig erkannt und gelöst."[11] Mit der Implementierung von geeigneten KPIs (Key Performance Indicator) können im Incident Management Effektivität und Effizienz der Prozesse gemessen werden. KPIs sind Leistungskennzahlen, mit denen Vorgänge in Unternehmen intern bewertet werden können und somit folglich ein Teil der internen Begutachtung sind.[12] Insbesondere sind KPI-Messungen für die Erstellung von Trends und Statistiken sehr nützlich, die neben der allgemeinen Kontrolle der IT-Services außerdem auch für die Motivation der Mitarbeiter gut verwendet werden können.[13] Neben der ökonomischen Säule wird hier somit noch der soziale Aspekt des Drei-Säulen-Modells der Nachhaltigkeit adressiert.[14] „Darüber hinaus ist die kontinuierliche Anwendung der vereinbarten KPIs ein wichtiges Erfolgskriterium. [...] Werden KPIs nicht nachhaltig angewendet, verfehlen diese ihren eigentlichen Zweck."[15] Solche KPI-Messwerte können z. B. mittels eines Dashboardsystems visuell dargestellt werden. Das Zusammenspiel der Steuerung ist in IT-Systemlandschaften im Bereich des IT-Lieferantenmanagements einzuordnen.[16] Lieferantenmanagement wird definiert als die systematische Bewertung, Selektion und Entwicklung von Lieferanten.[17] Mit einem Dashboard ist für diesen Fall ein dialog- und entscheidungsorientiertes Instrument zu verstehen, das aktuelle und entscheidungsrelevante interne und externe Informationen benutzerfreundlich anbietet. Neben der quantifizierenden liefert ein Dashboard auch eine qualifizierende Verdichtung, indem vordefinierte Analyse und benutzerdefinierte Alarmmeldungen installiert werden.[18]

2 Nachhaltige Transparenz in IT-Service Landschaften durch das AMS Dashboard

In der Forschungskooperation *Dynamic IT Supplier Management* mit der Carl von Ossietzky Universität Oldenburg und einem Unternehmen aus der Automotivbranche wurde ein solches Dashboardsystem, mit der Bezeichnung AMS Dashboard entwickelt, mit dem ein Mitarbeiter aus dem Application Service Management in Abhängigkeit von individueller Schwellenwertbestimmung Kennzahlen für die „Anzahl Incidents" generieren und visuell darstellen kann. Diese Kennzahlen können kontinuierlich angewandt werden, so dass nachhaltig der Unterschied zu vorigen Messwerten ersichtlich ist. Das AMS Dashboards bietet einem Mitarbeiter aus dem Application Service Management (ASM) die Möglichkeit, diese Kennzahlen individuell für die IT-Services seiner Abteilung zu generieren und die Berechnungen visuell abzurufen, um somit einen entsprechenden Status zu erhalten. Darüber hinaus hat der Abteilungsleiter die Möglichkeit sich eine Gesamtübersicht dieser Kennzahlen in Form eines aggregierten Status einzuholen, um sich so eine bessere Übersicht bzw. Kontrolle über die Abteilung zu verschaffen.

[11] [Le12, S. 83]
[12] Vgl. [Bl07, S. 83]
[13] Vgl. [WS14, S. 45-46]
[14] [Kr11, S. 63ff]
[15] [Ch09, S. 44]
[16] Vgl. [Li15]
[17] Vgl. [WS14, S. 231-244]
[18] Vgl. [Gr12, S. 625-634]

2.1 Anforderungen

Im Rahmen der Forschungskooperation wurden durch eine Expertenbefragung bei einem Unternehmen aus der Automotivbranche folgende Anforderungen an eine solche Kennzahl für die „Anzahl Incidents“ erhoben, in dem der Nutzer[19] die KPI für *Anzahl Incidents*[20] erstellen können soll:

1. Auswahl einer *Bezeichnung*
2. Auswahl einer *Beschreibung*
3. Auswahl einer *CI*[21] *ID*[22]
4. Auswahl, ob *niedrige Werte besser* sind
5. Auswahl von *Schwellenwerten* für eine Ampeldarstellung des Status Grün, Gelb und Rot ()
6. Auswahl einer *Diagrammart*
7. Auswahl eines *zeitlichen Rahmens*, in dem die Incidents untersucht werden sollen
8. Auswahl des *Ausgabeintervalls* (z. B. wochenweise Betrachtung)
9. Nutzer soll die erstellten KPIs auf einer Statusübersicht in zusammengefasster Form sehen können

Für die bessere Verständlichkeit werden diese Anforderungen zunächst einmal in ein UML Anwendungsfalldiagramm (in der UML Version 2.5) überführt und dann wird das daraus entwickelte Dashboardsystem aufgezeigt.

2.2 Anwendungsfall

Im folgenden Anwendungsfalldiagramm werden die essentiellen Aktionen der KPI Erstellung sowie des Aufrufs durch einen Nutzer dargestellt, die für die „Anzahl Incidents“ nötig sind und darauf basierend die zwei Hauptanwendungsfälle beschrieben.

[19] Application Service Manager (sowohl allgemeiner Mitarbeiter als auch Abteilungsleiter)
[20] Anzahl aller aktuell offenen Tickets (Störfälle)
[21] Vgl. [Eb08, S. 708]
[22] ID einer CI (Configuration Item), die sich auf einen IT-Service bezieht

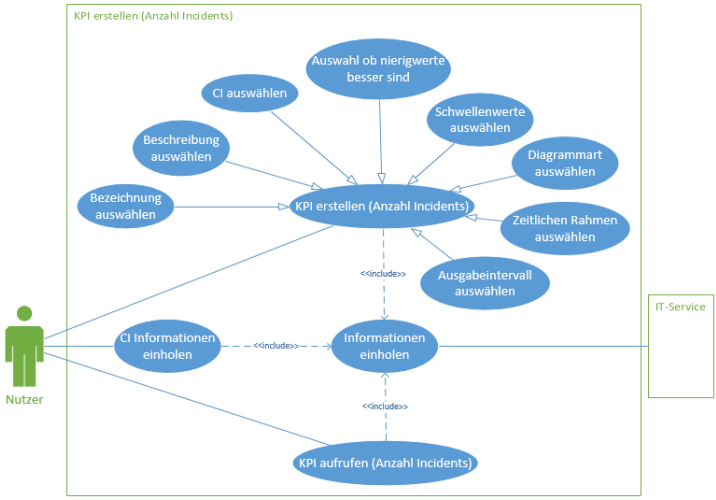


Abb. 1: KPI Anwendungsfall

UC1: (Nutzer erstellt Kennzahl)
Der Nutzer erstellt eine Kennzahl für die Anzahl Incidents, in dem er eine Bezeichnung angibt, eine dazugehörige Beschreibung, die CI ID auswählt, welche als Referenz für den entsprechenden IT-Service steht, die Angabe macht, ob niedrige Werte besser zu deuten sind, individuelle Schwellenwerte der Incidents bestimmt, einen zeitlichen Rahmen und den Ausgabeintervall auswählt.

UC2: (Nutzer ruft Kennzahl auf)
Der Nutzer ruft eine Kennzahl auf. Hierbei ist es möglich einzelne Kennzahlen auszuwählen und sich die entsprechenden Informationen einzuholen, wie die der Bezeichnung, die dazugehörige Beschreibung, die CI ID mit allen dazugehörigen Informationen des entsprechenden IT-Services, die Angabe, ob niedrige Werte besser zu deuten sind, die entsprechenden Schwellenwerte der Incidents, die er individuell bestimmt hat, den zeitlichen Rahmen, den Ausgabeintervall, den berechneten Status mit dem Report und der Darstellung in dem ausgewählten Diagramm. Neben der einzelnen Auswahl individuell erstellter Kennzahlen ist es auch möglich, über die Statusübersicht einen Gesamtüberblick aller individuell erstellten Kennzahlen in zusammengefasster Form zu sehen. In diesem Fall hat der Nutzer zusätzlich in der Abteilungsleiterrolle die Möglichkeit, sich eine Gesamtübersicht aller abteilungsbezogenen Kennzahlen in Form eines aggregierten Status einzuholen.

2.3 AMS Dashboard

Das AMS Dashboard wurde in Oracle APEX entwickelt, weil für die Umsetzung innerhalb der Forschungskooperation viele Vorteile von Oracle APEX eine große Rolle spielen, die

einen solchen Forschungszweck beschleunigen und systemunabhängig machen, da sie ein plattformunabhängiges Framework-System ist, eine schnelle Entwicklung von webbasierten Anwendungen mit direkter Anbindung zur Oracle Datenbank, einfacher Anbindung von anderen Datenquellen und die Nutzung der Sprache PL/SQL[23] bietet.
Die aus den Anforderungen resultierenden Angaben der Kennzahl können über die folgende Eingabemaske eingegeben werden, in dem die Bezeichnung, eine Beschreibung, die Auswahl des IT-Services, der Angabe, ob niedrige Werte besser sind, der Schwellenwertbestimmung, sowie der Diagrammart und der zeitlichen Eingrenzung mit ihrem Ausgabeintervall bestimmt werden. Für das folgende Beispiel wurden fiktive Daten verwendet:

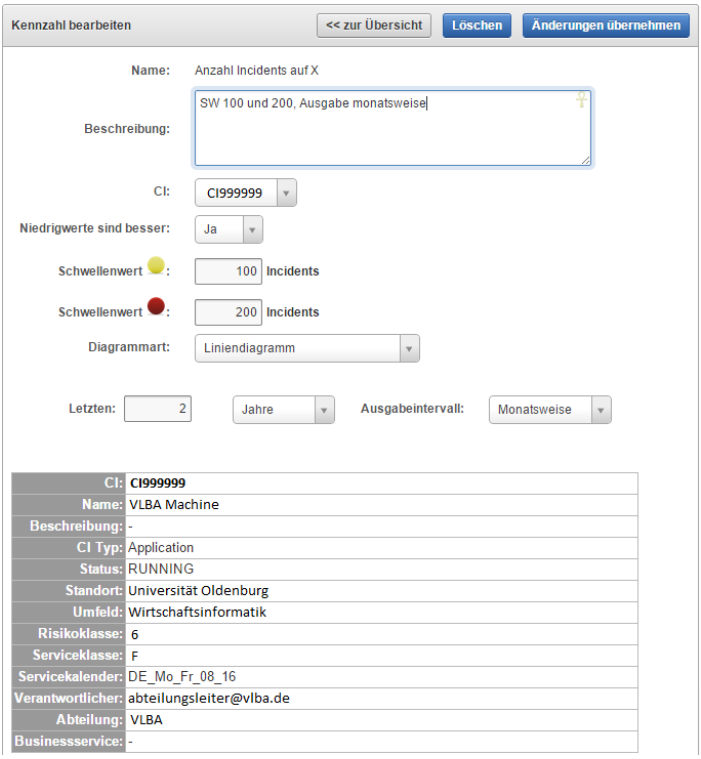


Abb. 2: KPI Anzahl Incidents (Angaben)

Nach Erstellung und dem Aufruf der berechneten Kennzahl wird die Ausgabe mit dem Report der Incidents, dem dazugehörigen Diagramm und dem Gesamtstatus der Kennzahl folgendermaßen wiedergegeben:

[23] PL/SQL (Procedural Language/Structured Query Language) ist eine Programmiersprache der Firma Oracle, die eine spezielle prozedurale Programmierung für SQL bietet

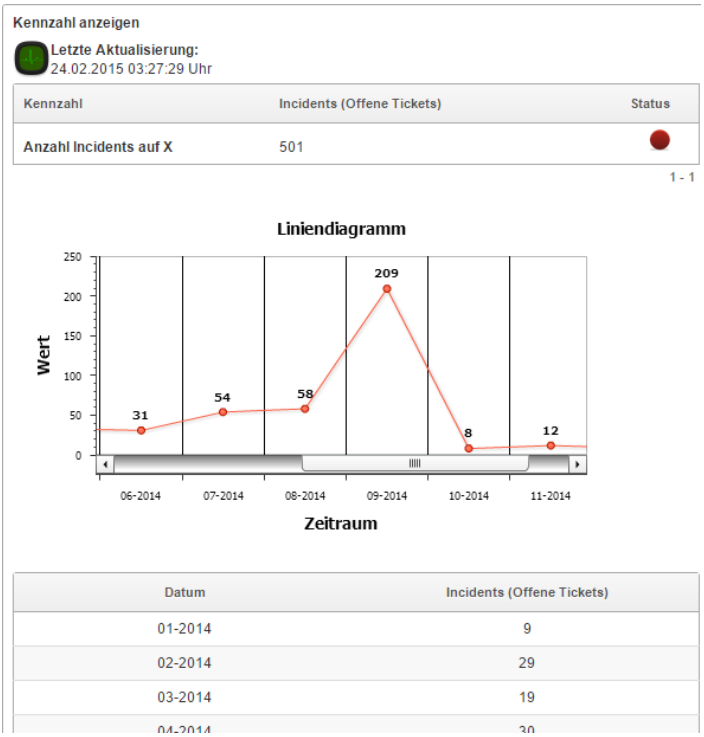


Abb. 3: KPI Anzahl Incidents (Visualisierung)

Durch die individuelle Bestimmung von Schwellenwerten der *Anzahl Incidents* auf den definierten Zeitraum soll eine Art Alarmierung ausgelöst werden. Je nach Status (Grün/-Gelb/Rot) hat der Nutzer die Möglichkeit in Abhängigkeit seiner Einschätzung entsprechend einzugreifen. So würde sich der Nutzer z. B. im Falle eines grünen Status keine Sorgen machen, wobei er im eintretenden Fall des gelben Status schon über Vorbeugungsmaßnahmen nachdenken würde und beim roten Status sofort handeln müsste.
Über die folgende Maske kann der Nutzer in einer zusammengefassten Statusübersicht alle von ihm erstellten Kennzahlen überblicken:

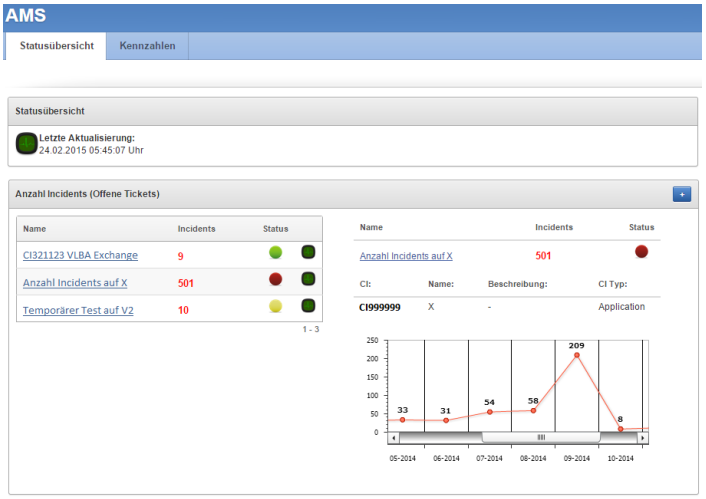


Abb. 4: KPI Statusübersicht Nutzer

Der Abteilungsleiter hat zusätzlich zur Erstellung der Kennzahlen die Möglichkeit sich einen Gesamtüberblick aller Kennzahlen auf einen Blick in einem aggregierten Status der Abteilungen zu verschaffen, dessen Berechnung als Durchschnitt aller Kennzahlen pro Abteilung erfolgt. Dieser Durchschnittsstatus wird als die Servicequalität in der Statusübersicht dargestellt, wie in folgender Abbildung zu erkennen ist:

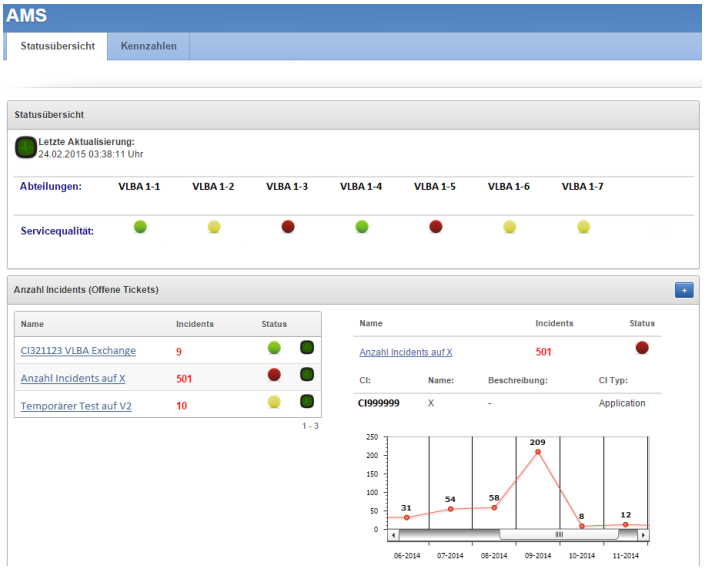


Abb. 5: KPI Statusübersicht Abteilungsleiter

3 Schlussfolgerung und Ausblick

Neben verschiedenen Dashboardsystemen auf dem Markt, die ähnliche Funktionalitäten bieten, wie in diesem Beitrag aufgezeigten AMS Dashboards, ermöglicht das AMS Dashboard insbesondere durch die Möglichkeit zur Priorisierung und Aggregation von Kennzahlen eine Transparenz im Incident Management, welche speziell mittels der Verkürzung von Alarmierungszeiten in kontextabhängiger Kennzahlengenerierung erlangt wird. Der Nutzer aus dem Application Service Management hat die Möglichkeit über die Definition individuell bestimmbarer Kennzahlen kontinuierlich die Anzahl von Incidents zu kontrollieren und sich eine nachhaltige Übersicht zu verschaffen. Das AMS Dashboard hilft dem Nutzer dabei einen ausführlichen Überblick der Informationsvielfalt zu bekommen und zusätzlich in zusammengefasster Form alle wichtigen Status auf einen Einblick einzusehen, um schnell in kritischen Situationen entsprechend intervenieren zu können. Für die kritische Einschätzung verhilft der Status, der mittels dreier Ampelfarben in Abhängigkeit der individuell bestimmten Kennzahlen alarmiert, um die nötige Handlungserfordernis einzuschätzen, womit die Steuerungsoptionen von IT-Services optimiert werden. Durch die Verkürzung der Alarmierungszeiten, welche in Abhängigkeit der individuellen Schwellenwertbestimmung erreicht wird, werden Zeiten eingespart, was folglich einen ökonomischen Vorteil in der Nachhaltigkeit bewirkt.
Das AMS Dashboard ist neben diesem Beispiel auch ohne Weiteres analog in anderen Bereichen des Incidents Managements nutzbar. Dafür ändert sich lediglich die Anbindung der Datenquelle. Das Verfahren jedoch bleibt gleich. Zum Beispiel kann dieses Konzept ebenso in der Nachhaltigkeitsberichterstattung mit diversen Umweltindikatoren als KPIs genutzt werden, um so ökonomische und soziale Verbesserungen zu bieten.
Um die Steuerungsmöglichkeiten mittels des AMS Dashboards neben der Transparenz allgemeiner Störfälle zu verbessern, würde ein Überblick der Reaktions- und Lösungszeiten im Zusammenhang von SLAs als weitere nützliche Information für dieses Lieferantenmanagementsystem verhelfen, mit dem sich die Forschungskooperation *Dynamic IT Supplier Management* mit der Carl von Ossietzky Universität Oldenburg und einem Unternehmen aus der Automotivbranche weiterhin beschäftigt.
Überdies ist es außerdem wichtig im Zusammenhang der immer größer werdenden Datenmengen, Maßnahmen der notwendigen Performanzoptimierung zu betrachten und den Forschungsschwerpunkt auf den Bereich Big Data zu fokussieren, um darüber hinaus den parallel entstehenden Vorteil der wachsenden Daten andererseits als Information effektiv zu nutzen.[24]

Literaturverzeichnis

[Ba13] Baron, P.: Big Data für IT-Entscheider: Riesige Datenmengen und moderne Technologien gewinnbringend nutzen. Carl Hanser Verlag GmbH & Company KG, 2013.

[Bl07] Bleiber, R.: Controlling für Nicht-Controller. Haufe-Praxisratgeber. Haufe Mediengruppe, 2007.

[24] Vgl. [Ba13, S. 1]

[Ch09] Christophel, D.: Key Performance Indicators zur Optimierung von Instandhaltungsprozessen. Igel, 2009.

[Eb08] Ebel, N.: Bd. 3. Pearson Deutschland, 2008.

[FBS12] Fröschle, H.P.; Burr, W.; Schrey, J.: Praktisches Service-Level-Management: Basiswissen, Implementierung, Praxisbeispiele. Symposion, 2012.

[GJO09] Gómez, J.C.M.; Junker, H.; Odebrecht, S.: IT-Controlling: Strategien, Werkzeuge, Praxis. Compliance digital. Erich Schmidt, 2009.

[Gr12] Grote, S.: Die Zukunft der Führung. SpringerLink: Bücher. Springer Berlin Heidelberg, 2012.

[Kr11] Kraus, P.: Die Auswirkung von Corporate Governance und Nachhaltigkeit auf den Unternehmenserfolg: eine Betrachtung im Kontext der wertorientierten Unternehmensführung. Eul, 2011.

[Le12] Leimeister, J.M.: Dienstleistungsengineering und -management. Springer Berlin Heidelberg, 2012.

[Li15] Lieferantenbewertung.de: , Lieferantenbewertung. http://lieferantenbewertung.de/theorie/th-index_prozess.html, 2005 (Letzter Zugriff am 08.03.2015).

[RK09] Rudolph, S.; Krcmar, H.: Servicebasierte Planung und Steuerung der IT-Infrastruktur im Mittelstand: Ein Modellansatz zur Struktur der IT-Leistungserbringung. Gabler Research. Gabler Verlag, 2009.

[Wi12] Wilkinson, J.: Bd. 3. Haren Publishing, Van, 2012.

[WS14] Wolf, K.; Sahling, S.: Incident Management: Komplexe Störungen in der IT erfolgreich beheben. Carl Hanser Verlag GmbH & Company KG, 2014.

[ZK13] Zarnekow, R.; Kolbe, L.: Green IT: Erkenntnisse und Best Practices aus Fallstudien. Springer Berlin Heidelberg, 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Enterprise Architecture as a Tool for Managing Corporate Social Responsibility

Potential Use of Enterprise Architecture for CSR Reporting

Bettina Horlach[1], Paul Drews[2] and Ingrid Schirmer[3]

Abstract: The need for acting sustainably is one of the most urgent, but challenging topics enterprises currently face. In order to implement corporate social responsibility (CSR) on each enterprise level, companies establish dedicated management structures. Notwithstanding that several tools are already available for managing CSR, there appears to be no tool which provides an integrated and transparent overview of CSR-relevant organizational units, processes and information systems within a company. In this paper, we propose enterprise architecture (EA) as a suitable tool to fill this gap and facilitate in turn CSR reporting. By drawing on the results of a literature review and an empirical study on CSR management, we first identify CSR management's information needs in this regard. In a second step, we map these needs using information provided by existing EA frameworks. In the outcome, we present a meta-model which provides a CSR-oriented EA view.

Keywords: Enterprise Architecture, Corporate Social Responsibility, Sustainability, Reporting

1 Introduction and motivation

In times of manifold global economic, ecological and social challenges due, inter alia, to an ongoing resource depletion and the speeding up of climatic change as result of industrialization and population growth, the need for a sense of sustainability is growing in society. Especially, the so-called "Generation Y" is actively seeking global sustainable development in private and professional life ([De09], [Pw08]). People urge enterprises to question their responsibility for creating a desired and sustainable world economy. Today, many companies introduce a specialized function within the company for managing the enterprise's corporate social responsibility [LR13]. Its task is to facilitate ecological, economic and social sustainability on the enterprise level.

There is a vast variety of methods, concepts and tools ([Ec07], [Ra11]) developed by academia and in practice to support CSR management in its daily operations. Analytical

[1] University of Hamburg, Department of Informatics, Vogt-Kölln-Str.30, 22527 Hamburg, horlach@informatik.uni-hamburg.de
[2] Leuphana University of Lüneburg, Institute of Electronic Business Processes, Scharnhorststr.1, 21335 Lüneburg, paul.drews@leuphana.de
[3] University of Hamburg, Department of Informatics, Vogt-Kölln-Str.30, 22527 Hamburg, schirmer@informatik.uni-hamburg.de

methods (e.g. ABC, cross impact and risk analysis), management systems (e.g. integrated management systems) [Ra11] and IT solutions for sustainability (e.g. Green IT/ IS) are already used by the CSR management ([De10], [Lo11], [Me10]) for sustainability planning, management as well as communication and reporting. However, none of the aforesaid tools provides an integrated overview of the structure and processes of the company and its ecosystem. Since CSR does not exist in isolation but is only working if it is integrated into the overall enterprise context and its ecosystem, such a holistic view is essential for implementing CSR successfully. Yet, CSR management can use such information to improve its CSR reporting and communication internally as well as externally by being able to clarify and visualize connections, dependencies and interrelations between CSR actions and its outcomes on the overall enterprise level and beyond to see the "big CSR picture". Planning and management as well as analyzing and controlling CSR management's activities and projects will be further facilitated.

EA with its multi-layered models of an organization [ARW08] can deliver such traceability and is therefore proposed in this paper as a potential tool for the CSR management. Based on the assumption that the CSR management is a stakeholder in terms of EA necessitating specific information, we propose a distinct architectural view on the enterprise and its ecosystem reflecting the information needs the CSR management has in order to fulfill its function. Thus, we explore how EA can help CSR management in fulfilling its information needs in this article.

2 Research Approach

In order to address the proposed research goal, we choose to follow the design science process by Peffers et al. [Pe07] as we seek to develop a new artifact that is able to solve the problem we identified in CSR practice (see section 4). For our motivation to research CSR and EA, we first compared the aims of the CSR and EA management disciplines in order to identify potential links between them [HP15]. The results show that there is ample room for integrating both disciplines. On this basis, we conducted a literature review in respect of CSR and EA in order to identify existing approaches linking these concepts [HP15]. In the third section, we present the results of this literature review.

Assuming that the CSR management is a stakeholder in terms of EA with a distinct view on the architecture, we consequently developed an EA meta-model describing CSR management's information needs following the approach by Aier et al. [Ai08]. Since the meta-model has to be adapted to the special concerns [Ai08], we collected those by conducting an empirical study with CSR management in the finance industry [HP15] and by considering additional literature. In the next step, we investigated the extent to which current EA frameworks already address these concerns by mapping the concerns with the existing information provided by the frameworks. The results of both steps are described in section 4. Based on the results, we finally propose an EA meta-model from the view of CSR management in section 5 in order to fill the information gaps.

3 Related Research/ Theoretical background

3.1 Corporate Social Responsibility management

Since enterprises are central actors in society, they are main contributors to the economic, ecological and social development [Me10]. Thus, they have a particular obligation for a society's sustainable development. Corporations have to ensure with appropriate management that the business operations do not harm the needs of the present and the future society and that they do not increase environmental problems [Wo87]. Recently, enterprises and research have recognized that an efficient CSR management is also an important component of business survival and success in the present and future ([KCW08], [Ec07], [Lo04]).

With new legal requirements like the EU directive for disclosure of non-financial and diversity information for large enterprises [Eu14] and under growing pressure from non-governmental organizations, the public and by employees, enterprises are nowadays more than ever forced to reveal their efforts and activities regarding the CSR management inside the organization and in relation to their ecosystem. In turn, enterprises mainly use their company's websites or specific CSR integrated reports for communicating such activities. With new standards like GRI 4 and the Integrated Reporting Framework by the IIRC, the extent of those reports have significantly increased. Companies no longer have to only report on more 'traditional' CSR topics like their approaches to environmental protection (biodiversity, waste, pollution) ([BE10], [Eu01], [In09], [LR13]) and employee engagement (health and safety, diversity, development of human capital, sustainable incentive schemes) ([BE10], [Eu01], [In09], [LR13], [Pe09]) inside the organization. A strategic CSR approach by integrating social, environmental, ethical, human rights and consumer concerns into the business model and the core strategy [In11] with creating a CSR-conform corporate governance ([In11], [Gl13]), CSR management systems, CSR performance controlling as well as sustainable knowledge management are now also crucial internal parts which need to be communicated. In addition, CSR and integrated reports are nowadays employed to address the need for disclosure of the sustainability of the value chain the enterprise is operating in. Extending the boundaries of the enterprise's responsibility with this approach, the ecological and social sustainability of the products [Gl13] along with the entire value chain and the improvement of this value chain's sustainability [In11] including environmental protection, human rights and business practices [Gl13] constitute further focal points for reporting.

In order to approach the challenge of CSR reporting, specialized reporting systems have been developed and have been available in the market for the past ten years [Kp12]. Mainly based on the reporting indicators provided by the GRI standards, these systems create the sustainability report by gathering and visualizing the existing information in a comprehensive manner based on predefined reporting limits and content. Therefore, a connection to manifold information systems such as, for example, environmental

management information systems (EMIS) for ecological information, to HR systems for social content and ERP systems are required as sources of information. With the new challenges arising with the Integrated Reporting Framework and GRI 4 by further extending the reporting scope to report on financial as well as dedicated external aspects and their interrelations, we propose that this approach by gathering isolated information is no longer sufficiently effective. Moreover, the visualization of the interdependencies of such information is nowadays crucial. Therefore, we propose to include the concept of enterprise architecture in sustainability reporting to clarify those complex coherences and to enable comprehensive and transparent sustainability reporting. Additionally, we suggest that EA can be applied in further CSR management and planning activities as well. In the following section, the concept of EA is explained.

3.2 Enterprise Architecture

Enterprises are highly complex socio-technical systems with manifold interrelations between their elements. In order to be able to manage this complexity, EA can be a helpful instrument due to its role of being *"a complete expression of the enterprise"* [Sc04]. EA consequently represents "*a master plan which acts as a collaboration force between aspects of business planning such as goals, visions, strategies and governance principles; aspects of business operations, processes and data; aspects of automation such as information systems and databases; and the enabling technological infrastructure of the business such as computers, operating systems and networks"* [Sc04]. By structuring the enterprise in different layers and by combining these layers with their respective elements, EA enables a holistic and transparent perspective on the enterprise. Pursuing the goal of business IT alignment, EA enables informed decision-making with regard to the complex interdependencies between processes, information systems and IT infrastructure. Furthermore, strategic alignment is pursued by embedding enterprise architecture insights in analytical and decision-making processes.

EA has a variety of business-related as well as IT-related stakeholders including senior management, the CIO, program managers and project managers ([Ha11], [VSV08]). Since some of the involved and affected stakeholders have conflicting interests, EA also acts as „*a medium to achieve a shared understanding and conceptualizing among all stakeholders involved and govern enterprise development based on this conceptualization*" [Op09]. Thus, enterprise architects have to find balanced solutions for conflicting problems [Ke12].

Various approaches and frameworks have been published for the structuring of an enterprise, its components and their various relationships ([Ma11], [Sc04]). While some of these approaches solely focus on the IT level and their alignment to the business without considering the business needs, the majority of frameworks enforces strategic alignment by taking the whole enterprise into account. Furthermore, a new stream of EA research proposes the extension of EA's scope to the enterprise environment including business partners, competitors and customers [DS14]. EA is consequently not only

useful for business IT alignment and strategic alignment, but it also – in line with its function as enterprise ecological adaptation – provides "*the means for organizational innovation and sustainability*" [La12] in the market. Thus, EA management also involves analyzing internal contradictions as well as identifying incoherencies in the relationships with external stakeholders [La12].

3.3 Existing approaches regarding the use of EA in CSR management

As stated in the introduction, CSR management is a potential stakeholder for EA management since it may use EA as an instrument for sustainability reporting as well as planning, management, analyzing and monitoring. A comparison of the goals of the EA management and the CSR management is underpinning this assumption [HP15]. For analyzing this assumption, we conducted a systematic literature review and analysis based on the methodology of vom Brocke et al. [Vo09]. With this step, we seek to investigate the coverage of this topic in the existing literature.

In the first iteration, we conducted a literature search focused on the interconnection of CSR and EA. We analyzed the leading journals on information systems (IS) of the AIS Senior Scholars' Basket Journals, the four leading IS conferences (AMCIS, ECIS, HICSS and ICIS) and several online literature databases and search engines (including *AIS electronic library, ACM Digital Library, EBSCOhost Online Research Databases, Emerald Insight, Google Scholar, IEEE Xplore Digital Library, JSTOR, ProQuest, SpringerLink, Wiley Online, wiso, and Web of Knowledge*). In a second iteration, we conducted a search in a set of leading CSR and business ethics journals and conferences (including *Business Ethics Quarterly, Business Strategy and the Environment, Corporate Social Responsibility & Environmental Management, Corporate Responsibility Management, Ethics and Information Technology, International Journal of Sustainable Strategic Management, International Journal of Value-Based Management, Journal of Business Ethics, Journal of Global Responsibility, Journal of Management and Governance and Social Responsibility Journal)* [Un11]. Finally, a structured google search has been conducted in order to find 'grey' literature and archival data produced by practice and academia.

Our literature review aimed at uncovering existing publications which explicitly discuss the realization of CSR in or with the aid of enterprise architecture. We used the general search term "(Corporate Social Responsibility OR Sustainability OR Sustainable Development) AND (Enterprise Architecture)" and searched the described list of journals, conferences and literature databases for articles containing the terms in the title, abstract, keywords or full text. For these publications, a forward and a backward search was conducted [WW02]. After scanning title, abstract and full text of the identified articles, we soon discovered that existing literature is predominantly either addressing the sustainability of EA itself based on flexibility, modularity [AD05] and adaptability to internal and external changes [Ma12]. Another set of literature promotes the planning and implementation of CSR with regard to EA inside the enterprise on two levels: either

solely focusing on sustainable IT or addressing the sustainability of the organization in total. The IT-centric position promotes sustainable "*enterprise IT architecting*" [La12] by employing EA for an optimized use of IT assets in a sustainable manner [Pa13]. This is realized by selecting and enforcing IT projects, mainly with regard to environmental topics ([No10], [Pa13], [Un11]), in the EA management process. These IT projects address on the one hand the idea of "Greening of IT" [PRS13] in the form of, for example, the introduction of less energy consuming hardware. On the other hand, these projects focus on using IT in order to enhance sustainability under the slogan of "Greening through IT" [PRS13] by e.g. enforcing video communication. The enterprise-centric literature enriches these thoughts by extending the use of EA for planning sustainable transformation on a functional level. EA management therefore does not only let ecological and social aspects influence its decision as to which IT projects are to be implemented in the future. The planning and decision-making in a sustainable manner also effects business-related initiatives. EA is in this regard to be envisioned as the "link between strategy development and execution" [Me10] by deciding which projects are to be implemented in order to reach the desired target architecture. These approaches thus view EA as a sustainability planning instrument on company level ([Ev11], [Me10], [No10], [SCC13]). Communication and reporting are however only mentioned by Scholtz [SCC13] in the context of environmental reporting as providing a source of information without going into depth.

Since the CSR management extends the external sustainability view seeking to foster the stakeholder dialogue and consequently transform the value chain in a sustainable manner, the scope of EA needs to cover the information about the enterprise environment as well. However, none of the researched articles is taking this external perspective into consideration with regard to CSR. Although e.g. Medini and Bourey [MB12] propose an approach for supply chain management in the context of EA and EA management, CSR related aspects have not been addressed. Furthermore, though this approach uses an external EA perspective by involving the supply chain into internal business processes and capabilities, the architectural representation of the stakeholders in the supply chain and their involvement remains unclear. For instance, information about the current supply chain structure is not integrated in the EA model. This limits the possibilities for analyzing the supply chain including its collaboration and working relationships with existing and future suppliers and business partners. The question, whether EA in the function of "*enterprise ecological adaptation*" [La12] could also be involved in identifying these aspects instead of solely representing current internal processes, applications and capabilities is therefore not answered in the literature. Since the CSR management however needs information on both the internal structure and the environment, we propose a design solution overcoming the shortcoming in the literature. By identifying CSR management's concerns first and consequently building the architectural representation of the information needed for realizing these concerns, we developed an extended EA meta-model representing both internal and external CSR related aspects. The process of development and the evolved meta-model are described in the following sections in extent.

4 Identifying CSR management's EA concerns

The initial step of the approach proposed by Aier et al. [Ai08] contains the elicitation of concerns by the respective stakeholders in order to identify the distinct information needs. In order to identify these needs, we conducted a qualitative-empirical study with 13 CSR managers in the German-speaking finance industry. We analyzed their scope of activity and the challenges they currently have to face. The detailed results of this empirical study will be published in [HP15]. Since enterprise architecture has no or only minor influence on some CSR concerns in the empirical study (e.g. a missing common definition of CSR or a lacking external pressure to pursue CSR inside the organization), we excluded these issues from the further analysis. The remaining topics were then analyzed for stakeholder concerns addressed by the interviewees. Based on the results, we identified 11 EA-relevant CSR management concerns (see table 1).

	Internal Perspective	**External Perspective**
Economic Sustainability	• CSR risk dashboard • Integrated reporting • CSR management system	• Big data usage for real-time CSR analyses • Sustainable investments
Ecological Sustainability	• Product lifecycle analyses • Digitalization of information • Printer reduction	• Supply chain analyses
Social Sustainability	• Product lifecycle analyses • Occupational safety app	• Supply chain analyses • Cloud solution for collaborative stakeholder dialogue

Tab. 1: Stakeholder concerns by CSR management

The concerns show that the use of IT is a focal point in CSR management. In the majority of the enterprises we interviewed, the management is using IT for their daily operations like measuring, reporting, stakeholder collaboration or cross-cutting management tasks like risk, product or supply chain management. CSR managers also see that the digitization of data for storage and analyses results in positive environmental effects. Initial CSR aspects like occupational safety, ecological action (e. g. reduction of printers to reduce emissions), and analyses on the supply chain are also still relevant topics for CSR management.

The current challenge addressed most often is the CSR reporting both in- an external with the goal of establishing an integrated report. Since this report aims at integrating sustainability aspects with the financial performance of the enterprise, it reveals how a company creates financial and non-financial value with its business activities and outputs [19]. The contribution of sustainable business habits on the enterprise's financial state is explained by showing their interrelations. Since such a report covers the overall

sustainability and financial performance, it influences and is highly being influenced by all other topics addressed by the interviewees. Integrated reporting can thus be seen as the overall CSR management frame. Since an integrated and holistic thinking in regard to the internal structure as well to the environment of the enterprise is especially important for successfully fulfilling this task, we propose that enterprise architecture can in particular be useful for the reporting internally and externally. Hence, we will further suggest how its information needs can be facilitated by EA.

5 An EA meta-model for supporting CSR management

5.1 Selecting appropriate meta-models for modelling CSR concerns

For investigating the extent to which the CSR reporting and the further concerns are already addressed by current EA frameworks, we analyzed five selected approaches by The Open Group [Th11], Krcmar [Kr10], Dern [De09], Winter and Fischer [WF06] and Hanschke [Ha11] if they meet the following set of criteria: Since CSR is a holistic integrated concept affecting both the business and the IT inside the organization as well as externally, an approach has to reflect this accordingly. All impacts of CSR strategies and activities on the enterprise shall be taken into account. Furthermore, the existence of a strategy layer in the approach is required since sustainability is directly influencing and occasionally transforming the enterprise strategy due to its integrating character [In11]. Finally, the approach has to provide architectural layers, architectural elements belonging to the respective layers as well as relations between these elements.

Based on these criteria, none of the approaches matches all required demands. Although the frameworks can represent the internal enterprise view in a comprehensive way, the external dimension is underrepresented in the approaches. Concerns requiring manifold external information like product lifecycle or supply chain management are not fully captured by the frameworks since crucial stakeholder information as well as relevant business architecture elements are missing. Nevertheless, their results are essential prerequisites for integrated reporting [In11] representing the current state of the company's internal and external CSR performance. Furthermore, we concluded that the corporate strategic approach of CSR, which is compulsory for the integrated CSR reporting, has to be extended as well since crucial strategic architectural elements for CSR like the corporate mission and vision are missing in the identified approaches. In contrast, the IT-related concerns like extending an IT-based enterprise risk dashboard to also include ecological and social risks for calculating their possibilities of occurrence and repercussions can already be realized by the analyzed approaches because they contain all relevant information on application systems, data and infrastructure.

Based on these results, we propose an EA meta-model from the view of CSR management overcoming these shortcomings. For constructing our meta-model we use

selected frameworks in order to use the experience/knowledge of the existing approaches. Therefore, Winter and Fischer's model [WF06] was selected since it fulfills most of the listed requirements except the missing external orientation and the non-existing representation of the relations between architectural elements. Since relations are however relevant for allowing the investigation of CSR measures and their impact on the sustainability performance, we additionally used TOGAF and its full content meta-model [Th11] with all proposed extensions. This meta-model represents the interrelations between architecture elements in a comprehensive way. The practice-orientation of TOGAF, being based on the experience of a manifold number of enterprise architecture practitioners, supports our decision for selecting this framework as a second source.

5.2 An EA metal-model for supporting CSR management

Based on our understanding of CSR, we created a meta-model that includes the information needed for supporting the CSR management's concerns (see figure 1). This meta-model uses the architectural layers and selected elements from Winter and Fischer's approach (e.g. organizational unit, business function, strategy) [WF06] as well as selected elements (e.g. business capability, business process, business service, data object) [Th11] and relations from the TOGAF content meta-model as described above. It has been extended by additional layers and elements, which support the CSR management's concerns. The extensions address three major areas: (1) additional elements in the "environment layer", (2) the business architecture and (3) the process architecture.

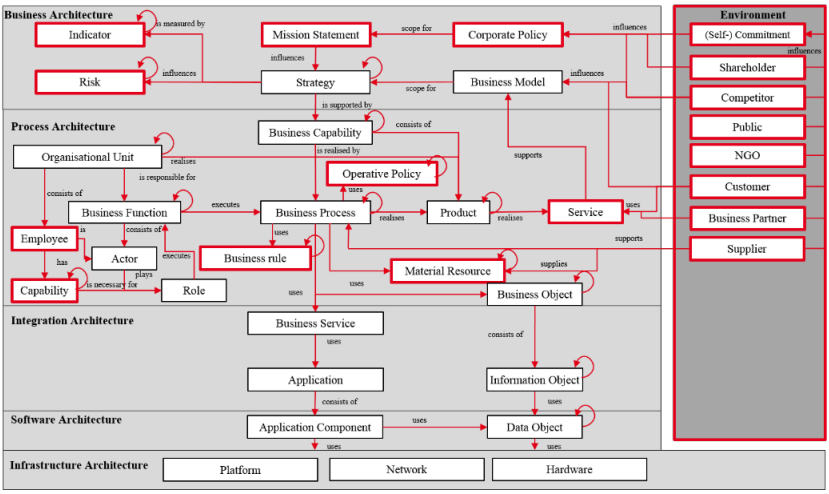


Fig. 2: Enterprise architecture meta-model for supporting CSR management

In contrast to the analyzed meta-models, our approach models the enterprise as an open socio-technical system, which is interrelated with its various **external stakeholders** [Rh13]. This external layer ("environment") follows the approach of integrated reporting since a major part of an integrated report contains the "insight into the nature and quality of the organization's relationships with its key stakeholders, including how and to what extent the organization understands, takes into account and responds to their legitimate needs and interests" [In11]. Since suppliers for instance provide resources like raw or prefabricated components to the enterprise, they directly influence the enterprise sustainability performance on the product level as well as with their business practices and behavior. Furthermore, the customers are a focal topic for reporting, as they influence the business model, processes and products with their habits and needs [Bi09]. The same is true for business partners as they act as distribution channels towards the customers. Finally, legislation and the public (including non-governmental organizations as "watchdogs" [Rh13]) have to be taken into account. They analyze the company's sustainability performance with regard to its behavior in general and seek information on compliance and trace self-imposed external and internal commitments. This leads to a new vertical layer, which we call "environment". This layer is only of preliminary nature and should be refined and split up to also consider other layers like processes or IT of the external actors in the future. Accordingly, the CSR management perspective supports the call for modeling the extended enterprise or business ecosystem architecture [DS14]. This step is e. g. required to keep track of impact from external actors and to integrate relevant information in complex supply chains.

Additionally, we propose to extend the **business architecture layer** towards **the strategic and normative dimension of CSR**. We argue that the consideration of the corporate strategy, business models, the normative elements (corporate policy, mission statement, corporate values) guiding the conduct towards employees and environment [BE10] is crucial for CSR management by acting as the starting point for all CSR activities. Analyses about the impact and conformance of specific actions cannot be conducted without determined policies and strategies. Furthermore, the integrated picture of the combination, interrelatedness and dependencies between strategy and the operational CSR actions is mandatory in the integrated report [In11]. Therefore, we propose to use Winter and Fischer's strategy layer for extending the architecture by adding the respective elements. The additional elements "risk" and "indicator" in this layer represent the effects of CSR on risk and cost management. Adding these elements fulfils the need of CSR reporting to identify the "risks and opportunities that affect the organization's ability to create value over the short, medium and long term" [In11]. However, risks and indicators have to be applied to all architectural layers in order to raise individual threats and to identify room for improvement.

In the final extension regarding CSR-relevant components, we propose to **enrich the process architecture layer**. This is necessary due to the important role of resources in the CSR context. The raw and prefabricated components provided by suppliers, the resulting finished product, but e. g. also human resources (employees and their capabilities) are seen as forms of capital for CSR management realizing the financial and

non-financial value for the organization [In11]. Here, the degree of a supplier's involvement in production processes is an essential information in case the supplier or his materials does not comply with mandatory supplier CSR policies or other obligations. In such a case, it is not contributing to creating value and consequently has to be replaced. By further separating the product resulting from executing a process and the final service, analyses can be accomplished in greater detail. But also the employees as social capital are a focal point to be analyzed and reported on from the CSR perspective as a contributor to service production and stakeholders of CSR. Therefore, the workforce is represented in our model by an actor with allocated roles and corresponding capabilities. Finally, normative elements like business rules and operational policies as the elements operationalizing the corporate strategy and normative values and guiding the business operations are also lacking. Since the business operations have to be compliant to the enterprise strategic direction [In11], information about the resources and their contributions to value delivery in form of services as well as the process organization have to be ascertained and modelled.

6 Conclusion and outlook

In this paper, we have argued that the use of enterprise architecture as an instrument for CSR management can support its operationalization. It allows to better understand the enterprise and its ecosystem and therefore helps to identify useful CSR activities and projects. Furthermore, it allows to improve the reporting of an enterprise's sustainability performance internally and externally by proposing the necessary information and their existing interrelations.

Our literature review and the comparison of five well-known existing architecture frameworks revealed that enterprise architecture approaches are insufficiently addressing the external dimension of CSR so far. Therefore, we compared information items derived from CSR management concerns with existing architecture frameworks. In the next step, we integrated architectural layers and elements based on the mapping into an architectural meta-model.

Our research is limited as the meta-model only gives a broad outline on how the application of enterprise architecture could support CSR management and CSR reporting in particular. Therefore, it has to be applied in a practical environment and needs to be enriched with additional concerns and information items. For future research, the criteria we used to enrich the existing meta-models therefore have to be extended in number and refined in detail. Additionally, the comparison of the information demand and the supply by existing enterprise architecture approaches has to further be extended as well in order to identify potentially suitable existing frameworks supplying CSR management's information needs.

7 References

[AD05] Aier, S.; Dogan, T.: Indikatoren zur Bewertung der Nachhaltigkeit von Unternehmensarchitekturen. 7. Internationale Tagung Wirtschaftsinformatik 2005, S. 607–626, 2005.

[Ai08] Aier, S. et al.: Stakeholderorientierte Dokumentation und Analyse der Unternehmensarchitektur. GI Jahrestagung, S. 559–565, 2008.

[ARW08] Aier, S.; Riege, C.; Winter; R.: Unternehmensarchitektur – Literaturüberblick und Stand der Praxis. Wirtschaftsinformatik. 50, S. 292–304, 2008.

[BE10] Baumgartner, R.J.; Ebner, D.: Corporate sustainability strategies: Sustainability profiles and maturity levels. Sustainable Development 18, S. 76–89, 2010.

[Bi09] Bickel, I.M.: Corporate Social Responsibility: Einflussfaktoren, Erfolgswirkungen und Einbezug in produktpolitische Entscheidungen. Diplomica Verlag, Hamburg, 2009.

[De10] Dedrick, J.: Green IS : Concepts and Issues for Information Systems Research. Communications of the Association for Information Systems, 27, 1, S. 174–183, 2010.

[De09] Deloitte: Finding the green in today's shoppers: Sustainability trends and new shopper insights, 2009, http://www.gmaonline.org/downloads/research-and-reports/greenshopper09.pdf, Stand: 05.04.2015.

[De09] Dern, G.: Management von IT-Architekturen: Leitlinien für die Ausrichtung, Planung und Gestaltung von Informationssystemen. 3., durchges. Auflage, Vieweg + Teubner, Wiesbaden, 2009.

[Dh14] D'Heur, M.: CSR und Value Chain Management: Profitables Wachstum durch nachhaltig gemeinsame Wertschöpfung. Springer Verlag, Berlin, Heidelberg, 2014.

[DS14] Drews, P.; Schirmer, I.: From Enterprise Architecture to Business Ecosystem Architecture. Trends in Enterprise Architecture Research (TEAR) 2014, 2014.

[Ec07] Econsense: Nachhaltigkeitsmanagement in Unternehmen - Von der Idee zur Praxis: Managementansätze zur Umsetzung von Corporate Social Responsibility und Corporate Sustainability, 2007, http://www.econsense.de/sites/all/files/nachhaltigkeitsmanagement_unternehmen.pdf, Stand: 05.04.2015

[Eu01] European Commission: Promoting a European framework for corporate social responsibility. 5, July, 2001, europa.eu/rapid/press-release_DOC-01-9_en.pdf, Stand: 30.03.2015.

[Eu11] European Commission: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A Renewed EU Strategy 2011-14 for Corporate Social Responsibility, 2011, http://ec.europa.eu/enterprise/policies/sustainable-business/files/csr/new-csr/act_en.pdf, Stand: 30.03.2015.

[Eu14] European Union: Directive of the European Parliament and of the Council amending Directive 2013/34/EU as regards disclosure of non-financial and diversity information by certain large undertakings and groups, http://register.consilium.europa.eu/doc/srv?l=EN&f=PE%2047%202014%20INIT, Stand: 10.04.2015.

[Ev11] Evernden, R.: Sustainable EA : Is It Possible ? Business Enterprise Architecture Executive Update 16, 9, S. 1–6, 2011.

[Gl13] Global Reporting Initiative: G4 Sustainability Reporting Guidelines – Reporting Principles and Standard Disclosures, 2013, https://www.globalreporting.org/resourcelibrary/GRIG4-Part1-Reporting-Principles-and-Standard-Disclosures.pdf, Stand: 10.04.2015.

[Ha11] Hanschke, I.: Enterprise Architecture Management - einfach und effektiv: Ein praktischer Leitfaden für die Einführung von EAM. Carl Hanser Verlag, München, 2011.

[HP15] Horlach, B.; Drews, P.: Unterstützung des Nachhaltigkeitsmanagements durch Unternehmensarchitekturmanagement – Ziele und Einsatzpotentiale. 12. Internationale Tagung Wirtschaftsinformatik. , Osnabrück (2015).

[In11] International Integrated Reporting Council: The international Integrated Reporting framework, 2011, http://www.theiirc.org/wp-content/uploads/2013/12/13-12-08-THE-INTERNATIONAL-IR-FRAMEWORK-2-1.pdf, Stand: 10.04.2015.

[In09] International Standard Organisation: ISO 26000 - Guidance on Social Responsibility, 2009, www.iso.org, Stand: 02.04.2015.

[KCW08] Kurusz, E.C.; Colbert, B.A.; Wheeler, D.: The Business Case for Corporate Social Responsibility. In: Crane, A. et al. (eds.) The Oxford Handbook on Corporate Social Responsibility. Oxford University Press, S. 83–112, 2008.

[Ke12] Keller, W.: IT-Unternehmensarchitektur: Von der Geschäftsstrategie zur optimalen IT-Unterstützung. 2., überarb. u. erw. Auflage, d.punkt Verlag, Heidelberg, 2012.

[Kp12] KPMG: Sustainability Reporting Systems – A market review. 2012, https://www.kpmg.com/DE/de/Documents/Sustainability-Reporting-Systems-2012-kpmg.pdf, Stand: 10.04.2015.

[Kr10] Krcmar, H.: Informationsmanagement. 5. Auflage, Springer Verlag, Berlin, Heidelberg, 2010.

[La12] Lapalme, J.: Three schools of thought on enterprise architecture. IT Professional 14, S. 37–43, 2012.

[Lo04] Loew, T. et al.: Bedeutung der internationalen CSR-Diskussion für Nachhaltigkeit und die sich daraus ergebenden Anforderungen an Unternehmen mit Fokus Berichterstattung: Endbericht an das Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit, 2004, http://www.upj.de/fileadmin/user_upload/MAIN-dateien/Themen/Einfuehrung/ioew_csr_diskussion_2004.pdf, Stand: 06.04.2015

[Lo11] Loos, P. et al.: Green IT: Ein Thema für die Wirtschaftsinformatik? Business Information Systems Engineering 53, S. 239–247, 2011.

[LR13] Loew, T.; Rohde, F.: CSR und Nachhaltigkeitsmanagement: Definitionen, Ansätze und organisatorische Umsetzung im Unternehmen. (2013), http://www.4sustainability.de/fileadmin/redakteur/bilder/Publikationen/Loew_Rohde_2013_CSR-und-Nachhaltigkeitsmanagement.pdf, Stand: 06.04.2015.

[Ma12] Magoulas, T. et al.: Sustainable Enterprise Architecture: A Three-Dimensional

Framework for Management of Architectural Change. European Conference on Information Management and Evaluation, 6, S. 178–186, 2012.

[Ma11] Matthes, D.: Enterprise Architecture Frameworks Kompendium - Über 50 Rahmenwerke für das IT-Management. Springer Verlag, Heidelberg, 2011.

[MB12] Medini, K.; Bourey, J.P.: SCOR-based enterprise architecture methodology. International Journal of Computer Integrated Manufacturing, 25, 7, S. 594–607, 2012.

[Me10] Melaragno, W.: Business Architecture : A Key Enabler for Sustainability Strategy Development and Implementation. STA Group Whitepaper, 2010.

[Me10] Melville, N.P.: Information Systems Innovation for Environmental Sustainability. MIS Quarterly 34, 1, S. 1–21, 2010.

[No10] Noran, O.: An Enterprise Architecture Approach towards Environmental Management Artefacts : A Brief Analysis. In: Bernus, P. et al. (eds.) Enterprise Architecture, Integration and Interoperability - IFIP TC 5 International Conference, EAI2N 2010, Springer Verlag, Berlin, S. 44–55, 2010.

[Op09] Op't Land, M. et al.: Enterprise Architecture - Creating Value by Informed Governance. Springer Verlag, Berlin, Heidelberg, 2009.

[Pa13] Pankowska, M.: Enterprise Architecture Modelling for Corporate Sustainability. International Conference on Information Systems Development, S. 365–376, 2013.

[Pe07] Peffers, K. et al.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24, 3, S. 45–77, 2007.

[Pe09] Peters, A.: Wege aus der Krise - CSR als strategisches Rüstzeug für die Zukunft Wege aus der Krise - CSR als strategisches Rüstzeug für die Zukunft, 2009, http://www.csr-weltweit.de/uploads/tx_jpdownloads/CSR-Trendstudie_Wege_aus_der_Krise_ final.pdf, Stand: 02.04.2015

[PRS13] Plessius, H.; Ravesteyn, P.; Slot, R.: Greening organizations : the Contribution of Enterprise Architecture. CARPE Conference, S. 1–12, 2013.

[Pw08] PwC International: Millennials at work Reshaping the workplace, 2008, http://www.pwc.de/de_DE/de/prozessoptimierung/assets/millennials-at-work-2011.pdf, Stand: 05.04.2015.

[Ra11] Raisiene, A.G.: Advantages and Limitations of Integrate Management System : Social Technologies 2011, 1, S. 25–36, 2011.

[Rh13] Rheinländer, K.: Vertrauen im betrieblichen Nachhaltigkeitsmanagement: Verhaltenstheoretische Betrachtung von Vertrauen als ökonomierelevante Kategorie. Kovac Verlag, Hamburg, 2013.

[Sc04] Schekkerman, J.: How to Survive in the Jungle of Enterprise Architecture Frameworks: Creating or Choosing an Enterprise Architecture Framework. 2. Auflage, Trafford Publishing, Victoria, 2004.

[SCC13] Scholtz, B.; Connolley, A.; Calitz, A: Enterprise Architectures for Addressing Sustainability Silos. International Conference on Information Technologies in Environmental Engineering, 6, S. 1–14, 2013.

[SH14] Sutherland, D.; Hovorka, D.: Enterprise Architecture as a Contributor to Research in Progress to Prior Research : Enterprise - IT alignment. European Conference on Information Systems, 22, S. 1–8, 2014.

[SS12] Schneider, A.; Schmidpeter, R.: Corporate Social Responsibility: Verantwortungsvolle Unternehmensführung in Theorie und Praxis. 2., erg. u. erw. Auflage, Springer Verlag, Berlin, Heidelberg, 2015.

[Th11] The Open Group: TOGAF 9.1, 2011, http://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html, Stand: 10.04.2015.

[Un11] Unhelkar, B.: Green Enterprise Architecture using Environmental Intelligence. International Journal of Green Computing, 2, S. 58–65, 2011.

[Un11] University of Southampton: CSR Journals, https://sotoncsrcommunity.wordpress.com/csr-journals/, 2011, Stand: 20.03.2015.

[Vo09] Vom Brocke, J. et al.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. European Conference on Information Systems, 17, S. 2206–2217, 2009.

[VSV08] Van Der Raadt, B.; Schouten, S.; van Vliet, H.: Stakeholder perception of enterprise architecture. Lecture Notes in Computer Science, 5292, S. 19–34, 2008.

[Wh06] Whitehouse, L.: Corporate social responsibility: Views from the frontline. Journal of Business Ethics, 63, S. 279–296, 2006.

[Wo87] World Commission on Environment and Development: Report of the World Commission on Environment and Development: Our Common Future (The Brundtland Report), 1987, http://www.un-documents.net/our-common-future.pdf, Stand: 30.03.2015

[WR06] Winter, R.; Fischer, R.: Essential layers, artifacts, and dependencies of enterprise architecture. IEEE International Enterprise Distributed Object Computing Conference Workshop, 10, S. 1–12, 2006.

[WW02] Webster, J.; Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 26, 2, S. 13–23, 2002.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Digitalisierung und IKT in der Nachhaltigkeitsberichterstattung von Hochschulen in Deutschland – terra incognita?

Ralf Isenmann[1], Sassen Remmer[2] und Sascha Zinn[3]

Abstract: Ziel des Beitrags ist es, den Einsatz von IKT in der Nachhaltigkeitsberichterstattung von Hochschulen in Deutschland als Element der Digitalisierung zu untersuchen. Dazu zählen z.B. Einzelanalysen zu Distributionskanälen und zur medialen Verfügbarkeit der Nachhaltigkeitsberichte, zu den Formen der IKT-gestützten Online-Kommunikation, die die Vermittlung der Inhalte erleichtern können, der Einsatz moderner IKT und ausgeklügelter Softwarewerkzeuge entlang der Kernprozesse der Nachhaltigkeitsberichterstattung - von der Erstellung der Berichte über die Verwaltung der Berichtsteile bis hin zur multimedialen Verteilung und anmutenden Präsentation - sowie zu Treibern, Hemmnissen und Erfolgsfaktoren in der Nachhaltigkeitsberichterstattung von Hochschulen anhand von ausgesuchten Fallbeispielen. Mit dem Beitrag soll eine Brücke geschlagen werden zwischen dem etablierten Forschungsbereich der Nachhaltigkeitsberichterstattung von Unternehmen und den neuerdings intensivierten Aktivitäten von Hochschulen in Deutschland auf der einen Seite und der betrieblichen Umwelt- und Wirtschaftsinformatik auf der anderen Seite. Der Bedarf an konzeptioneller Orientierung in der Forschung und an praktischer Handreichung in den Hochschulen ist groß.

Keywords: Digitalisierung, Hochschule, Internet, Interaktivität, Nachhaltigkeitsbericht, Nachhaltigkeitsberichterstattung; Software, Stakeholder-Dialog, Web 2.0

1 Einführung zur Digitalisierung und zum Einsatz von IKT in der Nachhaltigkeitsberichterstattung

Eine Vielzahl an Neuerungen gesetzlicher Vorschriften, Strukturvorgaben, Leitfäden und Empfehlungen erzeugen in der Nachhaltigkeitsberichterstattung eine hochdynamische Entwicklung [Is15a, Is13a; We11]: Die europäische Strategie zu Corporate Social Responsibility mit der EU-Richtlinie zur Offenlegung nichtfinanzieller und die Diversität betreffender Positionen durch bestimmte große Gesellschaften und Konzerne, die 4. Generation der Leitlinien der Global Reporting Initiative, der Ansatz zum Integrated Reporting sowie das "white paper" zum Einfluss von Internet-Technologien, insbesondere moderner Auszeichnungssprachen wie XBRL (eXtensible Business

[1] Hochschule München, Fakultät Betriebswirtschaft, Nachhaltiges Zukunftsmanagement, Am Stadtpark 20, 81243 München, ralf.isenmann@hm.edu
[2] Universität Hamburg, Institut für Wirtschaftsprüfung und Steuerwesen, Lehrstuhl für Revisions- und Treuhandwesen, Max-Brauer-Allee 60, 22765 Hamburg, Remmer.Sassen@wiso.uni-hamburg.de
[3] Hochschule München, Fakultät Betriebswirtschaft, Nachhaltiges Zukunftsmanagement, Am Stadtpark 20, 81243 München, sascha.zinn@hm.edu

Reporting Language), sind nur einige Beispiele für die Entwicklungen. Rasante Fortschritte in den modernen IKT überlagern und beschleunigen diese inhaltliche und formale Dynamik in der Nachhaltigkeitsberichterstattung selbst nochmals [Is15a; Is14; FT10].

So ermöglichen komfortable Publikationsumgebungen ein computergestütztes Publizieren. Groupware-Technologien unterstützen die arbeitsteiligen, oftmals funktionsbereichsübergreifenden Redaktionsprozesse in der Berichterstattung. Verfahren aus dem Document Engineering erleichtern die formalisierte Erstellung standardisierter Dokumentstrukturen für Berichte. Sie erlauben zudem ein effizientes Dokumentenmanagement sowie eine medienübergreifende Präsentation von Berichten, vor allem durch den Einsatz standardisierter Auszeichnungssprachen wie XML und spezifischer Anwendungen, z.B. als ebXML (Electronic Business using XML) und XBRL für fortgeschrittene Internet-Applikationen und standardisierten Datenaustausch zwischen berichterstattenden Organisationen und ihren Zielgruppen. Ferner sind Software-Werkzeuge zur Unterstützung der Datensammlung, -verarbeitung und (teil)automatisierten Berichterstellung für verschiedene Medien verfügbar [IMG07; BV12; KPMG12]. Eine leistungsfähige Breitbandkommunikation ermöglicht es, Dokumente und andere Berichtsinstrumente auch mit großem Datenvolumen zügig zu übertragen. Ferner verstärken die Herausforderungen der globalen „Internet-Ökonomie" sowie die aktuellen Bestrebungen zur „Digitalisierung" in Politik und Wirtschaft insgesamt den Einsatz von Internet-Anwendungen in Unternehmen.

Aus den aktuell sich abzeichnenden Tendenzen, die die Entwicklung in der Nachhaltigkeitsberichterstattung insgesamt prägen, erscheinen die Digitalisierung und der Einsatz von IKT als ein wesentliches und treibendes Moment [Is07]. Es sind vor allem vier zentrale Merkmale, an denen die Digitalisierung und der Einsatz von IKT in der Nachhaltigkeitsberichterstattung deutlich werden [IsMG08]:

- Verfügbarkeit von Nachhaltigkeitsberichten in Print-, elektronischen und computergestützten Medien,
- neue Formen der Online-Kommunikation, darunter: Zielgruppenorientierung, Interaktivität, Stakeholder-Dialog und Web-2.0-Funktionalität,
- IKT-Unterstützung entlang der Kernprozesse sowie
- Einsatz von Software-Werkzeugen.

Entlang aller oben genannter Merkmale bietet der Einsatz moderner IKT für die Nachhaltigkeitsberichterstattung viele gute Gründe [Is11; IBW07; Is04; Le03]. Die Potenziale zur Informatisierung, Automatisierung und Personalisierung entfalten sich als informatik-bezogene Leitvorstellungen gerade dann, wenn die Nachhaltigkeitsberichterstattung strategisch verankert, in Organisation und Personalbedarfsplanung eingebunden, im Informationsmanagement der berichterstattenden Akteure integriert und in Informationssysteme operativ eingebunden ist.

2 Digitalisierung und IKT in der Nachhaltigkeitsberichterstattung von Hochschulen

Mit der Einführung zur Digitalisierung und zum Einsatz von IKT in der Nachhaltigkeitsberichterstattung ist die Basis gelegt für detaillierte Untersuchungen an Hochschulen in Deutschland.

- Zunächst werden die Besonderheiten zur Nachhaltigkeitsberichterstattung an Hochschulen beschrieben, und Hochschulen werden als Akteure charakterisiert, die neuerdings Nachhaltigkeitsberichte erstellen (Abschnitt 2.1).
- Vor diesem Hintergrund werden sodann ausgewählte Aspekte zur Digitalisierung und zum Einsatz von IKT in der Nachhaltigkeitsberichterstattung von Hochschulen in Deutschland untersucht und im Lichte der Erfahrungen von Unternehmen interpretiert (Abschnitt 2.2).

2.1 Hochschulen als nachhaltigkeitsberichterstattende Akteure

In den zurückliegenden 20 Jahren, seit ca. Mitte 1990er Jahren, hat sich die Nachhaltigkeitsberichterstattung zu einem bedeutsamen Thema für Unternehmen und ihre Stakeholder, Wissenschaftler und standardsetzende Institutionen entwickelt, und zwar mit hoher Entwicklungsdynamik und zunehmender Relevanz [Wa08]. Die Forschung zur Nachhaltigkeitsberichterstattung von Unternehmen kann als umfassend, differenziert und etabliert gelten [Fi12; KPMG13; Ko10; speziell zu IKT: Is14; FT13; IMG08a]. Es liegen eine Fülle konzeptioneller Erkenntnisse und ein großer Schatz an praxisnahem Erfahrungswissen vor.

Im Gegensatz dazu steht die Forschung zur Nachhaltigkeitsberichterstattung von Hochschulen vergleichsweise am Anfang, sei sie konzeptionell oder eher empirisch ausgerichtet. Dies gilt generell auf internationaler Ebene [Al14; LJ14; Si13] sowie speziell auch für die Hochschullandschaft in Deutschland [SDB14; MC09]. Bis dato haben nur wenige Hochschulen in Deutschland ihre mögliche Rolle als nachhaltigkeitsberichterstattender Akteur wahrgenommen und einen eigenständigen Nachhaltigkeitsbericht veröffentlicht: Laut einer aktuellen empirischen Studie [SDB14] haben zum Stichtag 31.7.2014 in Deutschland insgesamt 14 Hochschulen – das entspricht rund 4 % - einen Nachhaltigkeitsbericht veröffentlicht. Die Tendenz in den vergangenen Jahren ist leicht steigend.

Auch wenn es sicherlich nachvollziehbare Argumente gibt, warum Hochschulen die dokumentgebundene Form der Nachhaltigkeitsberichte als „Flaggschiff-Instrument“ der Nachhaltigkeitskommunikation noch nicht en gros für sich nutzen, so sprechen doch einige Gründe dafür, dass in Zukunft auch Hochschulen vermehrt Nachhaltigkeitsberichterstattung betreiben:

- So unterliegen sie spätestens seit der Rio-Konferenz 1992 und der Agenda 21 zunehmend der Erwartung, einen besonderen Beitrag zu einer nachhaltigeren Entwicklung zu leisten.
- Gleichzeitig stellt sich immer mehr heraus, dass auch Hochschulen selbst nachhaltiger werden müssen.

Die UN-Dekade Bildung für Nachhaltige Entwicklung [DUK11] und das UN-Weltaktionsprogramm Bildung für Nachhaltige Entwicklung [DUK13a] fordern dies ausdrücklich auf internationalen Ebene. In Deutschland bekräftigen dies z.B. die nationale Nachhaltigkeitsstrategie der Bundesregierung und der Nationale Aktionsplan der BNE-Dekade für Deutschland [DUK11], die Erklärung der Hochschulrektorenkonferenz [HRK11] sowie in die verschiedenen thematischen Broschüren der Deutschen UNESCO-Kommission [DUK11, DUK13b, DUK14a, DUK14b], flankiert und unterstützt durch Aktivitäten des Rates für nachhaltige Entwicklung [RNE14] und des BMBF wie z.B. die Symposien „Sustainability in Sciences (SISI)“ in 2013 und 2014.

Die Entscheidung zur Nachhaltigkeitsberichterstattung mit der freiwilligen Veröffentlichung von Nachhaltigkeitsberichten bietet in der Regel den Anstoß für eine regelmäßige, systematische und umfassende Auseinandersetzung mit Nachhaltigkeit in allen Handlungsfeldern einer Institution. Damit wird deutlich, dass neben den Nachhaltigkeitsberichten als jeweilige Zwischenergebnisse mit Dokumentationsfunktion einer zurückliegenden Berichterstattungsperiode der Nutzen der Nachhaltigkeitsberichterstattung vor allem darin liegt, einen zukunftsorientierten und kontinuierlichen Prozess der systematischen Schwachstellenanalyse, der stetigen Verbesserung und des gemeinsamen Lernens in Richtung Nachhaltigkeit anzustoßen. Dieser Anstoß zu einem solchen institutionalisierten Prozess erfasst in idealer Weise die gesamte Hochschule in allen ihren Facetten, Akteuren, organisatorischen Einheiten und Handlungsfeldern. Insofern kann der Nachhaltigkeitsberichterstattung die Funktion einer „Speerspitze“ zukommen.

In Anlehnung an Müller-Christ et al. [MC09] können Hochschulen mit Nachhaltigkeitsberichten drei grundlegende Zwecke erfüllen:

- Sie können generell Rechenschaft ablegen gegenüber ihren verschiedenen Stakeholdern über die von der Gesellschaft zur Verfügung gestellten - auch finanziellen – Ressourcen und ihre erbrachten Leistungen.
- Sie können ihren spezifischen gesellschaftlichen Beitrag als Bildungsinstitution in Forschung, Lehre, Weiterbildung und Transfer für eine nachhaltige Entwicklung vermitteln (Außenperspektive).
- Sie können darlegen, wie sie sich als wirtschaftende Organisation und öffentlicher Arbeitgeber z.B. in Betriebsführung und im Campusmanagement im Detail selbst nachhaltig ausrichten (Innenperspektive).

Die hier idealtypische gedankliche Unterscheidung zwischen einer Außenperspektive mit den oben genannten traditionellen Handlungsfeldern einer Hochschule auf der einen Seite und einer Innenperspektive mit Betriebsführung, Campusmanagement und anderen Aktivitäten auf der anderen Seite ist auch für die strategische Ausrichtung von Nachhaltigkeitsberichten bedeutsam. Denn es gilt z.B. festzulegen, wo Hochschulen ihre schwerpunktmäßigen Handlungsfelder in punkto Nachhaltigkeit sehen:

- vor allem im Beitrag zu einer nachhaltigeren Gesellschaft durch Forschung, Lehre, Weiterbildung und Transfer
- und/oder in der internen Optimierung des wirtschaftenden sozialen Systems „Hochschule“ als Non-Profit-Einrichtung mit Betriebsführung, Campusmanagement und anderen Aktivitäten [Is13b].

Diese akzentuierende Unterscheidung erscheint deshalb wichtig, weil für die Nachhaltigkeitsberichterstattung speziell an Hochschulen noch keine spezifischen Strukturvorgaben bzw. Berichtsstandards vorliegen [BDS15]. Gleichwohl gibt es eine Fülle von Ansätzen zur Erfassung und Bewertung von Nachhaltigkeit an Hochschulen [FJT15; Is15b]. Im Ergebnis variieren die bislang veröffentlichten Nachhaltigkeitsberichte der Hochschulen beträchtlich. Viele der von Hochschulen herangezogenen Strukturvorgaben zur Nachhaltigkeitsberichterstattung zielen schwerpunktmäßig auf die Innenperspektive, so z.B. die Leitlinien der GRI [2013a, 2013b] und Der Deutsche Nachhaltigkeitskodex [RNE11, RNE12].

Zum einen entstehen aus dem Zustand der Freiwilligkeit und der noch nicht erreichten Standardisierung für Hochschulen kreative Freiräume, die sie konzeptionell und innovativ zur inhaltlichen und medialen Ausgestaltung nutzen können [MC09]. So zeigt denn auch die empirische Untersuchung für Deutschland [SDB14] mit den bislang veröffentlichten 21 Nachhaltigkeitsberichten an Hochschulen ein insgesamt doch heterogenes Bild. Zum anderen entsteht aber auch die Anforderung, einen harmonisierten Vorschlag zu den hochschulspezifischen Anforderungen für Nachhaltigkeitsberichte zu erarbeiten, z.B. ähnlich den branchenspezifische Besonderheiten, wie sie die GRI in so genannten „sector supplements“ vorgelegt hat. Ein solcher Vorschlag der Klärung, Vereinheitlichung und Standardisierung wird vermutlich umso dringlicher empfunden, je mehr Hochschulen Nachhaltigkeitsberichte erstellen und das Bedürfnis nach Vergleich und ggf. Auszeichnung, Benchmarking, Ranking und Rating aufkommen mag.

2.2 Aktueller Stand zu Digitalisierung und IKT in der Nachhaltigkeitsberichterstattung von Hochschulen in Deutschland

Digitalisierung und IKT werden große Gestaltungschancen für die Nachhaltigkeitsberichterstattung zugesprochen. Ob und in welcher Weise die Hochschulen in Deutschland davon aktuell Gebrauch machen und in welcher Weise sie die sich bietenden Potenziale bereits ausschöpfen, soll hier anhand von vier

ausgewählten Aspekten näher untersucht werden [En15]:

- Verfügbarkeit von Nachhaltigkeitsberichten in Print-, elektronischen und computergestützten Medien,
- neue Formen der Online-Kommunikation, die Zielgruppenorientierung, Interaktivität, Stakeholder-Dialog und Web-2.0-Funktionalität eröffnen,
- IKT-Unterstützung entlang der Kernprozesse sowie
- Einsatz von Software-Werkzeugen.

Methodische Grundlage der Untersuchung zur medialen Verfügbarkeit der Nachhaltigkeitsberichte sowie zum Angebot der Online-Kommunikation ist eine Analyse der Websites bzw. Internetauftritte der 14 nachhaltigkeitsberichterstattenden Hochschulen (Tab. 1). Bei den 14 Hochschulen wurde neben den aktuellen Nachhaltigkeitsberichten auch eigenständige thematische Bereiche im Internet zu „Nachhaltigkeit“ und verwandten Begriffen wie z.B.: „Sustainable Development“ oder „Nachhaltige Entwicklung“ einbezogen. Die empirische Analyse wurde zwischen 17.-20. März 2015 durchgeführt und Änderungen noch im Zeitraum vom 1.-4. Mai 2015 mit aufgenommen.

Hochschule	Dateiformate der Berichte			Hypertextualität	
	PDF	Excel	HTML	Intern	Extern
Universität Bayreuth	1	0	0	1	1
Universität Bremen	1	0	0	1	1
Universität Duisburg-Essen	1	0	0	1	1
Hochschule Eberswalde	1	0	0	1	1
KU Eichstätt-Ingolstadt	1	0	0	1	1
Universität Hamburg	1	0	0	1	1
Hochschule Heilbronn	1	0	0	1	1
Fachhochschule Kaiserslautern	1	0	0	0	1
Universität Lüneburg	1	0	0	1	1
Universität Oldenburg	1	0	0	1	1
Universität Osnabrück	1	0	0	1	1
Hochschule Trier	1	0	0	0	0
Universität Witten-Herdecke	1	0	0	0	0
Hochschule Zittau/Görlitz	1	0	0	1	1
Summe	14	0	0	11	12

Tab. 1: Mediale Verfügbarkeit der Nachhaltigkeitsberichte von Hochschulen in Deutschland

Alle 14 untersuchten Hochschulen stellen ihren Nachhaltigkeitsbericht neben der Printform auch als elektronisches Dokument zum Download (PDF) zur Verfügung. Gemäß einer herangezogenen dreiteiligen Einteilung von Berichtstypen zur Nutzung des Internets [IL02] handelt es sich hier zumeist um den Typ 1 „konvertierte Berichte“ sowie um Typ 2 „web-basierte Berichte. Die Einteilung in Typen kennzeichnet die

unterschiedlichen Ausprägungen der Nachhaltigkeitsberichte, so wie sie im Internet verfügbar sind. Sie geben Auskunft, wie das Internet für die Nachhaltigkeitsberichterstattung im Blick auf Dateiformate, technische Standards und in Abhängigkeit vom Grad der Ausschöpfung der medienspezifischen Unterstützungspotentiale genutzt wird. Drei Hochschulen, darunter Universität Lüneburg, Universität Oldenburg und Hochschule Zittau-Görlitz, bieten ergänzend eine spezifische Darbietung an, mit der sie eine „blätterbare Online-Ansicht" ermöglichen, nachempfunden dem Umschlagen von Seiten in Printwerken. In engem Zusammenhang mit dem Dateiformat steht die Erreichbarkeit des Nachhaltigkeitsberichts. Ein Maß der Erreichbarkeit ist die Anzahl der Klicks, um von der Startseite bzw. der Homepage der Hochschule zum Nachhaltigkeitsbericht zu gelangen. Bei vier Hochschulen sind die Nachhaltigkeitsbereiche direkt von der Startseite zu erreichen. Bei sechs Hochschulen sind bis zu drei Klicks nötig. Bei einer Hochschule sind fünf Klicks erforderlich, und in drei Fällen bedarf es der Suchfunktion, die Nachhaltigkeitsberichte zu erreichen.

Um die Formen der Online-Kommunikation zu erheben, wurde ein bestehender Kriterienkatalog [IF14] herangezogen und nochmals verfeinert. Der Kriterienkatalog basiert auf einer Konzeptualisierung von Web 2.0-Funktionalitäten einerseits [Sü12; SMI14] und einer Synopse empirischer Analysen zur Nachhaltigkeitsberichterstattung andererseits. Der Kriterienkatalog umfasst 30 Indikatoren, gebündelt in sechs Bereiche, darunter: Usability, Informationsangebot, Erstellung von Inhalten, Organisation und Strukturierung von Inhalten, Bewertung von Inhalten und Organisation der Nutzer (Tab. 2).

Kriterien	Indikator	Anzahl
Usability		
Erreichbarkeit von Startseite	Direkt von Startseite	4
	Max. drei Klicks	6
Sitemap		10
Suchfunktion	Suchfunktion im Nachhaltigkeitsbereich	4
Informationsangebot		
Berichtsform	Umweltbericht	1
	Nachhaltigkeitsbericht	11
	Integrierter Bericht	1
Dateiformate	PDF	14
	Excel	0
	HTML (online)	0
Hypertextualität	Intern	11
	Extern	12
Aktualität	Aktualisierung < drei Monate	5
Multimedia	Video	2
Erstellen von Inhalten		
	Beiträge verfassen, korrigieren oder erweitern	0
		0
Organisation und Strukturierung von Inhalten		

Tagging, Kategorisierung	Warenkorb-Funktionalität	0
	Layoutanpassung	0
	Navigation	1
	Lieblingslisten	0
	Nutzergenerierte Meta-Daten	0
Ähnlichkeitssuche	Recommender Engine	0
Verlinkung	RSS-Feed	2
Bewertung von Inhalten		
Kommentierung	Kommentierung des Berichts	0
	Kommentierung anderer Inhalte	4
Bewertung, Feedback	Kontaktmöglichkeit zu Nachhaltigkeitsbericht	12
	Online-Befragung	1
Organisation der Nutzer		
Kommunikation	Sharing per E-Mail	4
	Sharing in sozialen Netzwerken	5
Community, Gruppenbildung	Anlegen von Nutzerprofilen	0
	Folgen anderer Nutzer	0
Selbstorganisierter Dialog	Nachhaltigkeits-Blog	2

Tab. 2: Kommunikative Vermittlung der Nachhaltigkeitsberichte von Hochschulen in Deutschland, im Lichte von Web-2.0-Funktionalitäten

Über alle hier untersuchten Indikatoren hinweg werden einige Kernbefunde deutlich: Vier Hochschulen verweisen direkt von der Startseite auf ihre Nachhaltigkeitsbereiche, und bei sechs Hochschulen gelangen Nutzer mit bis zu drei Klicks zu den Nachhaltigkeitsberichten. Die Erreichbarkeit sei ein Indiz für die Bedeutung, die die Hochschule dem Nachhaltigkeitsbericht beimisst. Bei den Berichtsformen überwiegen die thematisch integrierten Formen gegenüber reinen Umweltberichten. Hypermedia (Hypertext + Multimedia) scheint für viele Hochschulen eine geeignete Option zu sein, um die Vermittlung der Nachhaltigkeitsinhalte medial anzureichern. Die Möglichkeiten, dass Nutzer Inhalte zur Nachhaltigkeitsberichte beisteuern können, werden durchgängig - noch - nicht angeboten. Von den sich bietenden Optionen zur Organisation und Strukturierung von Inhalten machen Hochschulen bislang keinen Gebrauch, mit Ausnahme der Verlinkung durch RSS-Feed. Nahezu alle Hochschule bieten für ihren Nachhaltigkeitsbericht einen direkten Kontakt an, i.d.R. eine Ansprechperson. Eine Kommentierung von Berichtsinhalten ist bislang nicht implementiert. Gleichwohl bieten zwei Hochschulen auf ihren Nachhaltigkeits-Blogs via Kommentarfunktion, dort Einträge zu veröffentlichen. Zwei Hochschulen bieten auf ihren Social-Media-Kanälen (Facebook), Kommentare einzufügen. Neun Hochschulen unterstützen die Organisation der Nutzer, darunter vier via Sharing per E-Mail und fünf via Sharing in sozialen Netzwerken. Hingegen spielt die Gruppenbildung der Nutzer in der Nachhaltigkeitsberichterstattung offensichtlich noch keine Rolle.

Bei einer einem Ranking nachempfundenen Auswertung, aber ohne „naming, shaming and blaming“, ergibt sich in punkto Nutzung von Web-2.0-Funktionalitäten in der

Nachhaltigkeitsberichterstattung eine Spitzengruppe von fünf Hochschulen [En15]: Universität Lüneburg, Hochschule Zittau/Görlitz, Universität Hamburg, Hochschule für nachhaltige Entwicklung Eberswalde und Hochschule Heilbronn sowie Universität Osnabrück und KU Eichstätt-Ingolstadt. Diese Hochschulen greifen im Vergleich zu den anderen die medienspezifischen Unterstützungspotenziale intensiver für ihre Nachhaltigkeitsberichterstattung auf und scheinen damit derzeit zu experimentieren, ganz im Sinne eines organisationalen Lern- und Entwicklungsprozesses. Gleichwohl bewegen sich Experimentierfreude und Grad der Umsetzung bislang noch auf einem einfachen Niveau.

Um die Erkenntnisse aus der oben skizzierten Bestandsaufnahme zu vertiefen, wurden zwischen Dezember 2014 und Februar 2015 aus der der oben genannten Gruppe in drei unterschiedlichen Hochschulen je ein ergänzendes Experteninterview zur Nachhaltigkeitsberichterstattung durchgeführt [En15; zur Methode: GL10; Mü81]. Die befragten Experten in den Hochschulen kommen aus den Bereichen: Nachhaltiges Management, Wirtschaftsinformatik sowie Umweltkoordination, und alle drei sind maßgeblich an der Erstellung der jeweiligen Nachhaltigkeitsberichte beteiligt, sei es redaktionell, technisch und/oder koordinativ. Sie gelten an der jeweiligen Hochschule als Schlüsselpersonen für die Nachhaltigkeitsberichterstattung. Die Experteninterviews liefern erste Hinweise zu Treibern und Hemmnisse liefern, wie die Prozesse zur Nachhaltigkeitsberichterstattung organisiert und implementiert und wie eine dauerhafte Prozessintegration ausgestaltet werden kann [En15]:

- Bei den Hemmnissen spielen erwartungsgemäß Engpässe bei Ressourcen eine große Rolle. So nennen die Befragten den großen zeitlichen Aufwand für die Erstellung der Berichtsinhalte und für deren Pflege sowie die zur Verfügung stehenden knappen finanziellen Mittel. Auch die spezifische Organisationsstruktur in Hochschulen scheint eine „Baustelle“ darzustellen. Die tendenziell flache Hierarchie, die große Autonomie für Fakultäten bzw. Fachbereiche sowie das Prinzip von Freiheit in Forschung & Lehre lassen die Hochschule als „unübersichtliche Institution“ erscheinen. Als begünstigend für die Nachhaltigkeitsberichterstattung hingegen werden definierte Prozesse der Nachhaltigkeitsberichterstattung sowie klare Verantwortlichkeiten der beteiligten Personen empfohlen. Fehlt der Nachhaltigkeitsberichterstattung die Verankerung in Strategie und mangelt es bei der Institutionalisierung in der Organisation, dann führt dies zu erhöhtem Koordinations- und Abstimmungsaufwand. Eine Top-down-Vorgehensweise, wie sie in Unternehmen angenommen wird, ist nach Auskunft der Befragten in Hochschulen nicht möglich.
- Bei den eine Nachhaltigkeitsberichterstattung treibenden Faktoren sind sich die Befragten in ihren Einschätzungen weitgehend einig: So sehen sie es übereinstimmend als hilfreich und förderlich an, die Nachhaltigkeitsberichterstattung an Hochschulen in vielfältiger Weise zu unterstützen. Dazu gehören z.B. Initiativen aus der Hochschullandschaft, die Einrichtung nationaler Beratungsstellen, geeignete Dienstleistungsangebote zur

Ausgestaltung von Nachhaltigkeitsberichten, das Ausloben von Leistungsanreizen z.B. in Form von Wettbewerben und Preisen. Eine besonders wichtige Bedeutung kommt vor allem der Legitimation auf der Ebene der Kultus- und Bildungsministerien, der Hochschulrektoren sowie anderer Multiplikatoren zu.

- Beim Einsatz von Software-Werkezugen, der Nutzung von IKT und der Ausschöpfung der medienspezifischen Unterstützungspotenziale differieren die Meinungen der Experten. Es scheint derzeit noch unklar, wie Digitalisierung und Einsatz von IKT bei Hochschulen angemessen ausgestaltet und in der Nachhaltigkeitsberichterstattung an Hochschulen umgesetzt werden kann. Angeblich liefern mögliche kritische Diskurse, negative Kommentare und andere Begleiteffekte der Online-Kommunikation keine Bedenken, die Nachhaltigkeitsberichterstattung digital zu öffnen. Vielmehr sehen die befragten Experten in Dialog, Feedback und Kommentaren durchweg den Chancencharakter. Eine Rolle spielen hingegen die verfügbaren Ressourcen, vor allem in punkto Zeit, Geld und Personalkapazität, die Hochschulen an der Implementierung ausgefeilter Software-Werkezuge und anderer Informationssysteme hindern. Die Nachhaltigkeitsberichterstattung ist an sich bereits ein komplexes Unterfangen unter den angespannten Ressourcenverhältnissen.

Die hier hervorgehobenen Details vermitteln beispielhaft, wie die konzeptionellen und empirischen Erkenntnisse zur Digitalisierung und zum Einsatz von IKT in der Nachhaltigkeitsberichterstattung mit den Befunden aus den Experteninterviews verknüpft und insofern mit der aktuellen Praxis in Hochschulen gespiegelt werden.

3 Fazit

Der Beitrag beleuchtet den Einsatz von IKT in der Nachhaltigkeitsberichterstattung von Hochschulen in Deutschland als Element der Digitalisierung. Er wirft ein aktuelles Licht auf die Frage, ob und in welcher Weise Hochschulen in Deutschland die Gestaltungschancen moderner IKT für die Nachhaltigkeitsberichterstattung einsetzen und die sich bietenden medienspezifische Potenziale ausschöpfen. Dazu wurden vier zentrale Merkmale genauer untersucht: Distributionskanäle und mediale Verfügbarkeit der Nachhaltigkeitsberichte, Formen der IKT-gestützten Online-Kommunikation, die Hochschulen zur Vermittlung der Inhalte nutzen, Nutzung moderner IKT und ausgeklügelter Softwarewerkzeuge entlang der Kernprozesse der Nachhaltigkeitsberichterstattung, reichend von der Erstellung der Berichte über die Verwaltung der Berichtsteile bis hin zur multimedialen Verteilung und anmutenden Präsentation sowie daraus abgeleitet erste Befunde zu Treibern, Hemmnissen sowie Erfolgsfaktoren in der Nachhaltigkeitsberichterstattung von Hochschulen.

Über alle Einzelbefunde hinweg ist der Bedarf an konzeptioneller Orientierung in der Forschung und an praktischer Handreichung in den Hochschulen zur Nachhaltigkeitsberichterstattung groß. Deshalb soll hier eine Brücke geschlagen werden

zwischen dem bereits etablierten Forschungsbereich der Nachhaltigkeitsberichterstattung von Unternehmen und den neuerdings intensivierten Aktivitäten von Hochschulen in Deutschland auf der einen Seite und der betrieblichen Umwelt- und Wirtschaftsinformatik auf der anderen Seite.

Ein solches Zusammenspiel eröffnet sicherlich einen anregenden Austausch, vor allem mit den betreffenden Akteuren in (Kultus-)Ministerien, Hochschulverwaltung, Präsidium bzw. Hochschulleitung und Nachhaltigkeitsbeauftragten. Obwohl die Forschungsstränge in beiden Bereichen bislang noch weitgehend getrennt verlaufen, so bieten die konzeptionellen und empirischen Befunde in der betrieblichen Nachhaltigkeitsberichterstattung und aus der betrieblichen Umwelt- und Wirtschaftsinformatik doch eine Fülle an Erkenntnissen und einen Schatz an Erfahrungswissen mit robusten Handreichungen, die es lohnen, für die Nachhaltigkeitsberichterstattung von Hochschulen - nach sorgfältiger Prüfung und in abgestufter Weise - ausgeschöpft zu werden, und zwar zum Nutzen aller Beteiligten: der Hochschulen als neue nachhaltigkeitsberichterstattende Akteure als auch der Stakeholder als die anvisierten Zielgruppen und Nutzer der Nachhaltigkeitsberichte.

Literaturverzeichnis

[Al14] Alonso-Almeida, del Mar M. et al. Diffusion of sustainability reporting in universities: Current situation and future perspectives. Journal for Cleaner Production. 2014 (im Erscheinen).

[BDS15] Beth, C; Dienes, D.; Sassen, R. : Integrated Reporting für Hochschulen auf Basis eines standardisierten Nachhaltigkeitsberichts. In (Freidank, C.; Müller, S.; Velte, P., Hrsg.): Integrated Reporting als neue Herausforderung für die unternehmerische Steuerung, Überwachung und Berichterstattung. Erich Schmidt, Berlin, S. 697-719, 2015.

[BV12] Brands & Value: Nachhaltigkeit managen. Softwaresysteme für das Nachhaltigkeitsmanagement - richtig auswählen und richtig implementieren. Oktober 2012. Bremen, www.brandsandvalues.com, Stand 18.01.2014.

[DUK11] Deutsche-UNESCO-Kommission e.V. (Hrsg.): UN-Dekade „Bildung für nachhaltige Entwicklung“ 2005–2014. Nationaler Aktionsplan für Deutschland. VAS, Bonn, 2011.

[DUK13a] Deutsche-UNESCO-Kommission e.V. (Hrsg.): Hochschulen für eine nachhaltige Entwicklung. Ideen zur Institutionalisierung und Implementierung. VAS, Bonn 2013a.

[DUK13b] Deutsche UNESCO-Kommission: Vorschlag für ein Weltaktionsprogramm „Bildung für nachhaltige Entwicklung“ als Folgeaktivität der UN-Dekade „Bildung für nachhaltige Entwicklung“ nach 2014. Vorläufige Arbeitsübersetzung der Deutschen UNESCO-Kommission. Version vom 31. Juli 2013, UNESCO-Exekutivrat, Board 192 EX/6, Paris, 2013b.

[DUK14a] Deutsche UNESCO-Kommission e.V. (Hrsg.): Hochschulen für eine nachhaltige Entwicklung. Nachhaltigkeit in Forschung, Lehre und Betrieb einer Hochschule. VAS, Berlin, 2014a.

[DUK14b] Deutsche UNESCO-Kommission e.V. (Hrsg.): Hochschulen für eine nachhaltige Entwicklung. Netzwerke fördern, Bewusstsein verbreiten. VAS, Bonn, 2014b.

[En15] Englisch, L.: Nachhaltigkeitsberichterstattung an deutschen Hochschulen - Web 2.0-Funktionalitäten als terra incognita? Fakultät Betriebswirtschaft, Hochschule München, 2015.

[Fi12] Fifka, M.: The development and state of research on social and environmental reporting in global comparison. Journal für Betriebswirtschaft 62(1), S. 45-84, 2012.

[FJT15] Fischer, D.; Jenssen, S.; Tappeser, V.: Getting an empirical hold of the sustainable university: a comparative analysis of evaluation frameworks across 12 contemporary sustainability assessment tools. Assessment & Evaluation in Higher Education http://dx.doi.org/10.1080/02602938.2015.1043234.

[FT10] Freundlieb, M.; Teuteberg, F.: Status Quo der internetbasierten Nachhaltigkeitsberichterstattung: Eine länderübergreifende Analyse der Nachhaltigkeitsberichte börsennotierter Unternehmen. In (Schumann, M. et al., Hrsg.): Multikonferenz Wirtschaftsinformatik 2010. Universitätsverlag, Göttingen, S. 1747-1759, 2010.

[FT13] Freundlieb, M; Teuteberg, F: Corporate social responsibilty - a transnational analysis of online corporate social responsibility reports by market-listed companies: contents and their evolution. International Journal of Innovation and Sustainable Development, 7(1), S. 1-26, 2013.

[GL10] Gläser, J., Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse. Verlag für Sozialwissenschaften, Wiesbaden, 2. Aufl., 2010.

[GRI13a] Global Reporting Initiative (GRI): G4 Leitlinien zur Nachhaltigkeitsberichterstattung. Berichterstattungsangaben und Standardangaben. GRI, Amsterdam, 2013a.

[GRI13b] Global Reporting Initiative (GRI): G4 Leitlinien zur Nachhaltigkeitsberichterstattung. Umsetzungsanleitung. GRI, Amsterdam, 2013b.

[HRK11] Hochschulrektorenkonferenz (HRK), Deutsche UNESCO-Kommission (DUK): Hochschulen für eine nachhaltige Entwicklung. Erklärung der Hochschulrektorenkonferenz (HRK) und der Deutsche UNESCO Kommission (DUK) zur Hochschulbildung für nachhaltige Entwicklung – Ein Beitrag zur UN-Dekade „Bildung für nachhaltige Entwicklung“ vom 22.01.2010, 2011.

[IBW07] Isenmann, R.; Bey, C.; Welter, M.: Online reporting for sustainability issues. Business Strategy and the Environment 16(7), 487-501, 2007.

[IF14] Isenmann, R.; Frühholz, A. T.: Nachhaltigkeitsberichterstattung 2.0 – Web-2.0-Funktionalitäten in der Nachhaltigkeitsberichterstattung der DAX 30-Unternehmen. In (Wohlgemuth, V.; Marx Gómez, J.; Lang; C., Hrsg.): Konzepte, Anwendungen und Entwicklungstendenzen von betrieblichen Umweltinformationssystemen (BUIS). Shaker Aachen, S. 217-226, 2014.

[IMG08a] Isenmann, R. Marx Gómez, J. (Hrsg.): Internetbasierte Nachhaltigkeitsberichterstattung. Maßgeschneiderte Stakeholder-Kommunikation mit IT. Erich Schmidt, Berlin, 1. Aufl., 2008a.

[IMG08b] Isenmann, R.; Marx Gómez, J.: Software-Werkzeuge für CSR. Effizientes

Informationsmanagement und maßgeschneiderte Stakeholder-Kommunikation. Forum nachhaltig Wirtschaften (FNW). Das Entscheider-Magazin, (1), S. 94-95, 2008b.

[Is04] Isenmann, R.: Internet-based sustainability reporting. International Journal of Environment and Sustainable Development (IJESD) 3(2): 145-167, 2004.

[Is07] Isenmann, R.: Online-Nachhaltigkeitsberichterstattung: Trends, Gestaltungschancen, Befunde, Fallstricke. In (Michelsen, G.; Godemann, J., Hrsg.): Handbuch Nachhaltigkeitskommunikation. Oekom, München: 2. Auflage, S. 623-634.

[Is11] Isenmann, R.: Corporate sustainability reporting – how to benefit from ICT for communicating à la carte. In (Quaddus, M. A.; Siddique, M. A. B. (Eds.): The Handbook of corporate sustainability: Frameworks, strategies and tools. Edward Elgar, Cheltenham (UK), Northampton (USA), S. 207-248, 2011.

[Is13a] Isenmann, R.: The future of sustainability reporting. Institutional infrastructure and dynamics in the field. In (Page, B. et al., Hrsg.): EnviroInfo 2013. Environmental Informatics and Renewable Energies, 27th International Conference on Informatics for Environmental Protection - Informatics for Environmental Protection, Sustainable Development and Risk Management - Part I and II. Shaker, Aachen, S. 481-483.

[Is13b] Isenmann, R.: Eine Professur für eine nachhaltige Hochschule. Interview mit Prof. Dr. Ralf Isenmann, Inhaber der Professur für nachhaltiges Zukunftsmanagement an der Hoch-schule München. Hochschulen für eine nachhaltige Entwicklung. Ideen zur Institutionalisierung und Implementierung. Deutsche-UNESCO-Kommission. VAS, Bonn, S. 22-23.

[Is14] Isenmann, R.: Die Rolle des Internets für die Nachhaltigkeitsberichterstattung – vom Distributionskanal zum zentralen Einstiegsknoten. In (Fifka, M., Hrsg.): Nachhaltigkeits- und CSR-Berichterstattung verstehen und erfolgreich umsetzen. Springer-Gabler, Berlin, Heidelberg, S. 107-123.

[Is15a] Isenmann, R.: Zur Zukunft der Nachhaltigkeitsberichterstattung. Neue Wege der Dokumentation und Kommunikation. Business Reporting. Zeitschrift für Berichterstattung und Dokumentation 1(4), S. 112-116, 2015.

[Is15b] Isenmann, R: Erfassung und Bewertung für Nachhaltigkeit an Hochschulen – eine internationale Übersicht bestehender Systeme. Expertenkreis zur Anpassung des Nachhaltigkeitskodex für Hochschulen, Rat für nachhaltige Entwicklung (RNE). Berlin, 13.05.2015.

[Ko10] Kolk, A.: Trajectories of sustainability reporting by MNCs. Journal of World Business 45, 367-374, 2010.

[KPMG12] KPMG (2012). Systainability reporting systems. A market review. May 2012, www.kpmg.nl, Stand: 18.01.2014.

[KPMG13] KPMG, UNEP, GRI, Unit for Corporate Governance in Africa: Carrots and sticks. Sustainability reporting policies worldwide - today's best practice, tomorrow's trends. KPMG Advisory, Global Reporting Initiative, Unit for Corporate Governance in Africa, 2013.

[Le03] Lenz, C: Empfängerorientierte Unternehmenskommunikation. Einsatz der Internet-Technologie am Beispiel der Umweltberichterstattung. Eul, Köln, 2003.

[Lo11] Lozano, R.: The state of sustainability reporting in universities. The International Journal of Sustainability in Higher Education 12: 67-78, 2011.

[Lo14] Lopatta, K.; Jeaschke, R.: Sustainability reporting at German and Austrian universities. International Journal of Education Economics and Development, 2014 (im Erscheinen).

[MC09] Müller-Christ, G.; Isenmann, R.; Dembski, N.: Nachhaltigkeitsberichterstattung von Universitäten. In (Baumgartner, R.; Biedermann, H.; Zwainz, M Hrsg.): Öko-Effizienz. Konzept, Anwendungen und Best Practice. Hampp, München, S. 83-100, 2009.

[Mü81] Mühlfeld, C. et al.: Auswertungsprobleme offener Interviews. Soziale Welt 32(3), S. 325-352, 1981.

[RNE11] Rat für Nachhaltige Entwicklung (RNE): Der Deutsche Nachhaltigkeitskodex, www.nachhaltigkeitsrat.de, Stand: 25.03.2015.

[RNE12] Rat für Nachhaltige Entwicklung (RNE): Der Deutsche Nachhhaltigkeitskodex (DNK). Empfehlungen des RNE und Dokumentation des Multistakeholderforums am 26.09.2011. Berlin, www.nachhaltigkeitsrat.de, Stand: 25.03.2015.

[RNE14] Rat für Nachhaltige Entwicklung (RNE): Ergebniszusammenfassung der RNE-Bildungskonferenz „Vom Piloten zum Standard: Nachhaltigkeit in Forschung, Lehre und Betrieb implementieren“, www.nachhaltigkeitsrat.de, Stand 25.03.2015.

[SDB14] Sassen, R.; Dienes, D.; Beth, C.: Nachhaltigkeitsberichterstattung deutscher Hochschulen. Zeitschrift für Umweltpolitik & Umweltrecht (37), S. 258-277, 2014.

[SMI09] Süpke, D., Marx Gómez, J., Isenmann, R.: Web 2.0 driven sustainability reporting. In (Lytras, M. D.; Pablos de P. O. (Hrsg.): Social web evolution: Integrating semantic applications and web 2.0 technologies. IGI Global, Hershey (USA) et al., S. 31-41, 2009.

[SSP13] Siboni, B.; Sordo, C del; Pazzi, S.: Sustainability reporting in state universities. An investigation of Italian pionieering practices. International Journal of Social Ecology and Sustainable Development 4(2), S. 1-15, 2013.

[Sü12] Süpke, D.: Referenzarchitektur zur dialogbasierten Nachhaltigkeitsberichterstattung im Web 2.0. Shaker, Aachen, 2012.

[Wa08] Waddock, S.: Building a new institutional infrastructure for corporate responsibility. The Academy of Management Perspectives 22(3), S. 87–108, 2008.

[We11] Wensen, K. van et al.: The state of play in sustainability reporting in the EU. Publication commissioned under the European Union's Programme for employment and social solidarity – progress (2007-2013). Amsterdam, Berlin, 2011.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Active Sourcing, Matching und Nachhaltigkeitskommunikation beim eRecruiting

Andreas Mladenow[1], Christine Strauss[1]

Abstract: In den letzten Jahren hat der Einsatz von IKT-gestützter Personalbeschaffung die Kommunikation zwischen potenziellen Arbeitgebern und Bewerbern entscheidend verändert. Für Unternehmen bieten sich vielfältige Möglichkeiten einerseits eigene Inhalte online zu verbreiten, um sich als attraktiver Arbeitgeber zu präsentieren, und andererseits einen Talentepool aufzubauen. Während konventionelle Methoden der Rekrutierung zeit- und kostenintensiv sind, können Human-Ressource-Verantwortliche durch aktive Nutzung von Web 2.0-Technologien neue Talente und potenzielle Bewerber rasch, einfach, ortsunabhängig und zielgruppenorientiert erreichen. Neben der Eigendarstellung (etwa im Bereich der Nachhaltigkeitsberichterstattung) geben Tools wertvolle Hilfestellung bei Personalplanung, Personalbedarfsdeckung, Arbeitsmarktbeobachtung, Stellenbeschreibung, sowie Auswahl-, Einstellungs- und Evaluierungsprozessen. Aus Sicht des Kandidaten kann sich entsprechende Nachhaltigkeitskommunikation des Unternehmens beim Matching zu einem entscheidungsrelevanten Kriterium entwickeln. Für das Themenfeld Nachhaltigkeit im Human Ressource Management eröffnen sich zudem neue Chancen hinsichtlich Verfügbarkeit, Erreichbarkeit und Disponibilität von Personal. Dabei zeigt die vorliegende Untersuchung, dass Anwendungen wie Active Sourcing nicht immer zielführend sind, sondern in Abhängigkeit von der zu besetzenden Position, der Branche, den Stellenanforderungen und der Unternehmensgröße eingesetzt werden sollen. Der vorliegende Beitrag analysiert - basierend auf fünf teilstrukturierten Experteninterviews mit Human Ressource Managern von nachhaltigkeitsberichterstattenden Konzernen - den Nutzen von eRecruiting (z.B. durch Erreichbarkeit junger Zielgruppen, gesteigerter Effizienz, Informationsbeschaffung). Ferner werden in diesem Kontext potenzielle zukünftige Einsatzbereiche, Chancen und Herausforderungen dargelegt.

Keywords: eRecruiting, Mobile Recruiting, Zielgruppenorientierung, Active Sourcing, HRM, Personal, Employer Branding, Matching, Geschäftsmodelle, Stakeholder, Web 2.0-Funktionalitäten, Personalsysteme, Nachhaltigkeit, Nachhaltigkeitskommunikation, Nachhaltigkeitsberichterstattung.

1 Einleitung

Das Human Ressource Management (HRM) vieler Unternehmen hat sich im Rahmen des informations- und kommunikationstechnologie-gestützten Wandels der letzten Jahre nachhaltig verändert [Ec14]. Insbesondere beim Rekrutierungsprozess werden die neu entstandenen Kommunikationskanäle genutzt, um im sogenannte „war for talents“

[1] Universität Wien, Fakultät für Wirtschaftswissenschaften, Fachbereich eBusiness, Oskar-Morgenstern-Pl. 1, 1090 Wien, {andreas.mladenow, christine.strauss}@univie.ac.at

geeignetes Personal zu akquirieren, welches im Informationszeitalter als knappe und wichtige Ressource und als langfristige Investition für den Unternehmenserfolg angesehen wird [Ec14, Bh08, Gh14, St01, St91, Ml11].

In diesem Kontext ist vor allem das HRM von Konzernen und größeren Unternehmen stets bemüht den ständigen Kontakt mit Interessenten und möglichen Bewerbern durch den geeigneten Einsatz von Web 2.0 Applikationen aufrecht zu halten und Kontinuität in Kommunikation zu pflegen [Bo05, Di07]. Gerade im Bereich des Rekrutierungsprozesses bieten sich für die Unternehmen neue Möglichkeiten der Ansprache, um die anvisierte Zielgruppe zu erreichen [Ec14]. Die Maßnahmen im Zuge des sogenannten eRecruiting reichen von der aktiven Suche nach neuen Mitarbeitern („high potentials“) durch „active sourcing“, über „matching“-Methoden durch Online-Intermediäre bis hin zum Aufbau einer Marke auf sozialen Medien als attraktiver Arbeitgeber.

Im Zuge der Unternehmensdarstellung wird in diesem Zusammenhang vom HRM – je nach anvisierten Talenten – auch auf Nachhaltigkeitskommunikation gesetzt. Die unternehmensweite Nachhaltigkeit bezweckt dabei nicht allein die Reputation des Unternehmens zu erhöhen [Ec14], sondern wird mittlerweile auch als langfristig relevanter Strategiebestandteil beim HRM wahrgenommen und in wissenschaftlichen Beiträgen diskutiert [Gu14, Eh14, Co12, Ba12].

Vor diesem Hintergrund analysiert der vorliegende Beitrag die aktuellen Potenziale des eRecruiting für Konzerne bzw. größeren Unternehmen. Der vorliegende Beitrag ist in vier Kapitel gegliedert. Das folgende Kapitel 2 beinhaltet theoretische Grundlagen zur Rekrutierungsthematik für den HRM-Bereich mit Fokus auf eRecruiting. Kapitel 3 widmet sich zunächst der Auswertung einer empirischen Untersuchung. Basierend auf fünf teilstrukturierten Experteninterview werden in diesem Zusammenhang HRM-Experten von internationalen Unternehmen nach den Nutzeffekten zu eRecruiting befragt. Im zweiten Teil von Kapitel 3 wird auf Online-Intermediäre eingegangen, auf welche das HRM zunehmend zurückgreifen. Danach schließt der Beitrag mit einer Conclusio und einem kurzen Ausblick.

2 Theoretische Grundlagen

2.1 Teilprozesse der Personalrekrutierung

Personalrekrutierung beinhaltet Maßnahmen zur Suche, Auswahl und Einstellung von Mitarbeitern und bezweckt die Gewinnung einer hinreichenden Mitarbeiterzahl, die mit geeigneter Qualifikation zum benötigten Zeitpunkt, am richtigen Ort eingesetzt werden können [Ba12]. In diesem Zusammenhang zeigt Tab. 1 Teilprozesse der Prozessrekrutierung und die dabei anfallenden Maßnahmen aus Unternehmenssicht

[Ba12, Cl13, Me13].

Personalrekrutierung	Teilprozesse
Bedarf	Personalplanung durchführen (strategisch, operativ) Stellenbeschreibungen, Anforderungsprofile erstellen
Marketing	„Signaling“: Darstellung als attraktiver Arbeitgeber „Screening“: Beobachtung des aktuellen Arbeitsmarktes
Bedarfsdeckung	Methoden zur Personalbedarfsdeckung Maßnahmen planen und durchführen
Personalauswahl	Auswahlverfahren und Teilnehmerkreis festlegen Auswahlverfahren durchführen
Personaleinstellung	Verträge abschließen „Onboarding“: Einstellungen begleiten
Evaluierung	Rekrutierungsprozess evaluieren Verbesserungsmaßnahmen entwickeln

Tab. 1: Teilprozesse der Personalrekrutierung

Bei der Ermittlung des Personalbedarfs steht zunächst die Klassifizierung der zu besetzenden Stelle im Vordergrund. Aus dieser Stellenbeschreibung wird in weiterer Folge ein Anforderungsprofil erstellt, welches die Grundlage für die weitere Personalsuche bildet. Umso genauer und klarer dieses Profil ist, umso effizienter kann der weitere Rekrutierungsprozess ablaufen [Ba12].

Die Attraktivität eines Unternehmens als Arbeitgeber spielt eine große Rolle bei Qualität und Quantität der Bewerber für eine ausgeschriebene Stelle. Gerade bei der Suche nach hochqualifizierten Mitarbeitern, den sogenannten „high potentials“, ist es für ein Unternehmen essentiell, sich schon vor der tatsächlichen Personalsuche als attraktiver Arbeitgeber zu positionieren. Um dies zu erreichen, gibt es im klassischen Recruiting eine Vielzahl von Instrumenten bzw. Richtlinien, um ein möglichst positives Image bzw. positive Perzeption bei potenziellen Bewerbern zu erreichen [Ba12].

Dies kann am Beispiel der Hochschulabsolventen veranschaulicht werden. Unternehmen, die vorrangig diese Zielgruppe rekrutieren, versuchen mit diesen bereits während der Studienphase in Kontakt zu treten. Je früher bei Studierenden ein positives Bild eines Unternehmens aufgebaut werden kann, umso leichter kann später ein positives Arbeitgeberimage vermittelt werden. Um das zu erreichen, können Unternehmen an den Universitäten und Hochschulen Präsenz zeigen, indem sie

Vortragsveranstaltungen mitgestalten oder indem sie als Sponsor bei universitären Veranstaltungen auftreten oder auch durch Infostände bei Messen oder Veranstaltungen vertreten sind. Eine andere Möglichkeit ist die Vergabe von Praktikumsplätzen an ambitionierte Studierende oder die Abfassung von Masterarbeiten in Zusammenarbeit mit dem Unternehmen zu ermöglichen. Auch die Ausschreibung für Stipendien für besonders leistungsstarke Studierende schafft ein positives Image als späterer Arbeitgeber. Zur Imagepflege für die Unternehmen zählt beispielsweise auch die Teilnahme an Jobmessen.

Weitere Teilprozesse der Rekrutierung beinhalten Bedarfsdeckung, Auswahl, Einstellung und Evaluierung [Ba12]. Bei der Personalbedarfsdeckung wird zwischen interner (Besetzung durch firmeneigenen Mitarbeiter) und externer Bedarfsdeckung unterschieden. Im Zuge des sogenannten Onboarding, der Einstellungsbegleitung des rekrutierten Mitarbeiters, werden zudem begleitende Maßnahmen getroffen (vgl. Kap 2.3).

2.2 Einsatz von eRecruiting

Die traditionelle Form des Recruiting wurde ungefähr um die Jahrtausendwende durch Stellenanzeigen im Internet erweitert. Die Veränderung der letzten Jahre im Recruiting-Bereich betreffen neben den Tätigkeiten, Systemen und Workflows insbesondere die Nutzung neuer Tools und Kommunikationskanäle. Dies wird auf externe Entwicklungen und den veränderten Gegebenheiten am Markt (z.B. steigende Bewerberzahl, Globalisierung des Jobmarkts, verändertes Kommunikationsverhalten) zurückgeführt [Ec14].

Im Rahmen des eRecruiting wird die Personalbeschaffung heute durch Web 2.0 Werkzeuge ergänzt, wodurch der gesamte Prozess aktiver gestaltet wird. Dabei suchen Recruiter gezielt nach potenziellen Kandidaten und intensivieren die Kommunikation. So können vom HRM neue Talente rasch, einfach, ortsunabhängig und zielgruppenorientiert erreicht werden [Gh14, Me13]. Zudem geben HRM-Tools wertvolle Hilfestellung bei den weiteren Teilprozessen, wie Personalplanung, Personalbedarfsdeckung, Arbeitsmarktbeobachtung, Stellenbeschreibung, sowie Auswahl-, Einstellungs- und Evaluierungsprozessen. Neben dem Einsatz von konzern-eigenen Lösungen und Software-as-a-Service-Produkten wird auch auf Open Source Software (z.B. apptivo, HRtrack, WebHR) zugegriffen.

Die derzeit verwendeten Methoden reichen von Online Assessment Center über Online Bewerbungsgespräche bis hin zu einer datenbankunterstützten Erstellung von Anforderungsprofilen. Hinzu kommt die aktive Einbeziehung einer Vielzahl von verschiedenen Social Media Plattformen in das Employer Branding [Sc09b], aber auch in die interne und externe Mitarbeitersuche. Auf diese Weise wird versucht, eine neue Generation an gut ausgebildeten Arbeitnehmern, welche in bestimmten Branchen und

Nischen in nur geringerer Anzahl vorhanden sind, erfolgreich zu erreichen und für das eigene Unternehmen zu gewinnen.

Um in Zeiten von sozialen Netzwerken eine positive Arbeitgebermarke zu schaffen, wird es für Unternehmen immer wichtiger, aktiv als attraktiver Arbeitgeber im Web 2.0 aufzutreten. Gerade für die Bereiche Arbeitsmarktkommunikation, interne Kommunikation und Networking eignen sich soziale Netzwerke besonders gut. Allerdings kann die Teilnahme an diversen eRecruiting Maßnahmen auch dazu führen, dass sich viele weniger geeignete Kandidaten melden, die zusätzlichen Aufwand der Abwicklung verursachen [Di07].

Um die öffentliche Wahrnehmung zu schärfen, ist es wichtig eine klare Abgrenzung zu anderen Unternehmen zu schaffen und ein individuelles Markenversprechen zu entwickeln. Das externe Employer Branding bezieht sich auf die Aspekte Arbeitsmarktkommunikation, Networking, Bewerbermanagement (Umgang mit Bewerbern) und Corporate Reputation [Sc09b]. In diesem Zusammenhang spielt auch die Nachhaltigkeitskommunikation zwischen Unternehmen und ihren Zielgruppen eine immer wichtigere Rolle [Bh08, Ba12].

2.3 Nachhaltigkeitskommunikation beim Recruiting

Im Zuge einer neuen EU-Richtlinie werden ab 2017 etwa 6.000 Großunternehmen und Konzerne in der EU dazu verpflichtet über Ihre Nachhaltigkeitsperformanz zu berichten [Re15]. Von dieser Richtlinie sind Unternehmen mit mehr als 500 Mitarbeiter betroffen, die lt. Richtlinie 2014/95/EU nicht-finanzielle Informationen (inklusive der Informationen zu Umwelt-, Sozial-, Arbeitnehmer- und Arbeitnehmerbelangen, Korruptionsbekämpfung und Achtung der Menschenrechte) offenlegen müssen [Na15]. Die Nachhaltigkeitsberichterstattung zielt auf die Kommunikationen zwischen Organisationen und ihre Zielgruppen ab und verbindet drei bisher isolierte Berichterstattungen, nämlich Geschäfts-, Umwelt-, Sozialberichterstattung. Sie macht auf diese Weise auf Wechselwirkungen zwischen ökonomischer, ökologischer und sozialer Rolle der Organisation in der Gesellschaft aufmerksam [Is15].

Eine unternehmensweite Nachhaltigkeit geht heute über die ledigliche Einhaltung staatlicher Vorschriften hinaus, sie wird vielmehr zum Strategiebestandteil beim Recrutingprozess [Bh08, Bo05] und wird zur Imagepflege großer Unternehmen herangezogen [Co12, Eh14]. Turban und Cable [Tu03] argumentieren, dass Organisationen mit guter Reputation von potentiellen Bewerbern bevorzugt werden. Zudem spielen die eigenen Wertvorstellungen der potenziellen Bewerber bei der Auswahl eines potenziellen Arbeitgebers eine Rolle [Ch05, Hi02]; dementsprechend wird argumentiert, dass Unternehmen, die eine hohe Reputation hinsichtlich Nachhaltigkeit aufweisen, von gut ausgebildeten Kandidaten bevorzugt werden [Sc09a, Ba12]. Nachhaltigkeitskommunikation wird dementsprechend gezielt vom HRM

eingesetzt. Die genannten Studien weisen somit auf den Sachverhalt hin, dass Interessierte und potenzielle Bewerber sich verstärkt mit der ökologischen Rolle (welchen Beitrag leistet das Unternehmen zum Umweltschutz?) und der sozialen Rolle (welchen sozialen oder sozialpolitischen Beitrag leistet das Unternehmen?) des Unternehmens in der Gesellschaft beschäftigen. Durch die Anwendung der sogenannten „Global reporting incentive reporting guidelines“ kann untersucht werden, wie viel Aufmerksamkeit der Nachhaltigkeit im Unternehmen gewidmet wird [Gr15].

Tab 2 enthält ausgewählte „best-practice“-Beispiele, die Alternativen darstellen, wie der HRM-Bereich Nachhaltigkeitskommunikation beim Onboarding implementieren kann [Ba12].

“best practice”-Prozesse beim Onboarding	Implementierung von Nachhaltigkeitskommunikation
• Grundlegende Informationen vor dem ersten Tag am Arbeitsplatz • Nutzung eines formalisierten Orientierungsprogramms • Entwicklung eines verschriftlichten Onboarding Plans.	Berücksichtigung von Nachhaltigkeit bei der Einstellungsbegleitung: • Erwartungen als Teil des Prozesses • Inhalte als Teil des Programms • Ziele im verschriftlichen Onboarding Plan
• Gestaltung des Onboardings als partizipativen Vorgang • Monitoring des Onboarding Programms im Zeitablauf	• Berücksichtigung von Programmen wie das “WalMart’s Personal Sustainability Project”, bei dem Mitarbeiter in die Festlegung der Ziele eingebunden werden. • Mit neuen MitarbeiterInnen in Verbindung bleiben, um sicher zu stellen, dass die Nachhaltigkeit während des Onboarding Prozesses nicht verloren geht
• Einsatz entsprechender Technologien zur Unterstützung des Onboarding Prozesses	• Verzicht auf ausgedruckte Dokumente als Umweltbeitrag

Tab. 2: „Best practice“-Nachhaltigkeitskommunikation beim Onboarding

3 Einsatz von eRecruiting in der Praxis

Aufschluss über den aktuellen Einsatz von eRecruiting bei Großunternehmen und Konzernen, die bereits eine Nachhaltigkeitsberichterstattung durchführen, gibt die Auswertung von Interviews (Kap. 3.1) und die angeführten Online-Praxisbeispiele (Kap 3.2), die von den Großunternehmen aktiv beim eRecruiting genutzt werden.

3.1 Empirische Untersuchung

Im Folgenden werden die Ergebnisse einer empirischen Untersuchung [Ja15] interpretiert und diskutiert. Die Untersuchung basierte auf fünf teilstrukturierten Interviews, die nach der dokumentarischen Methode nach Nohl [No08] durchgeführt und ausgewertet wurden. Die fünf Interviewpartner waren HRM-Vertreter großer internationaler Unternehmen am Standort Wien.

Die befragten Großunternehmen nutzen verschiedenste Kommunikationskanäle, um das Image des eigenen Unternehmens gegenüber potenziellen Bewerbern zu erhöhen und unternehmensbezogene Inhalte zielgruppengerecht zu kommunizieren. Zu diesem Zweck wird die eigene Homepage („Karriereseite“) als wichtige Informationsplattform betrachtet. Für die Verbreitung von Stellenanzeigen werden die sozialen Medien Facebook, LinkedIn und XING verwendet. Dort gepostete Stellenanzeigen werden mit einem Link versehen, der auf die eigene Hompepage des Unternehmenes führt; ferner werden soziale Netzwerke wie XING und LinkedIn genannt, um als Unternehmen permanent als institutionelles Mitglied vertreten zu sein. Für das Employer Branding werden ferner die Plattform *Facebook* und der Microblogging Dienst *Twitter* genutzt, vereinzelt wird auch auf *Watchado* zurückgegriffen. Die Anwendung unterschiedlicher Kommunikationskanäle für das Employer Branding zielt darauf ab, den potenziellen KandidatInnen Einblicke in den Arbeitsalltag zu gewähren. Als Zielgruppe wird dabei insbesondere die junge Generation angesprochen, die Reputation bzw. die Brand Visibility des eigenen Unternehmens wird damit erhöht und der Dialog mit Interessenten und potenziellen Bewerbern gefördert. Zudem wird verstärkt Wert auf Kontakt mit der Zielgruppe gelegt, die authentische Einblicke in das Unternehmen erhalten soll; außerdem werden passende Kommunikationskanäle für die jeweilige Zielgruppe (z.B. IT-Experten, Hochschulabsolventen) gewählt und das Zeigen der eigenen Unternehmenskultur hervorgehoben. Aus Sicht des Kandidaten kann sich entsprechende Nachhaltigkeitskommunikation des Unternehmens beim Matching zu einem entscheidungsrelevanten Kriterium entwickeln.

Active Sourcing wird im Rahmen des eRecruiting von Großunternehmen und Konzernen eingesetzt. Die Nutzung hängt allerdings von der zu besetzenden Position ab, sowie von der Branche, den Stellenanforderungen und der Unternehmensgröße; als nützlich wird es bei größeren Projekten und spezialisierten Stellen empfunden. Bedenken werden jedoch bezüglich der Reputation als Arbeitgeber durch Active Sourcing angesprochen. Der Einsatz von Online Assessment Tools wird jedoch von den befragten Unternehmen in

Österreich als nicht lohnenswert erachtet und wird nur von einem einzigen der insgesamt fünf HRM-Vertreter im Zuge einer Selbsteinschätzung von Bewerbern im Rahmen eines Online-Formulars verwendet. Hierbei erscheint in der Wahrnehmung der HRM-Vertreter der Fokus auf persönliche Gespräche zielführender.

Insgesamt wird das eRecruiting von allen fünf befragten Unternehmensvertretern positiv bewertet. Dabei wird einerseits besonders der Nutzen durch die gezielte Erreichbarkeit junger Zielgruppen positiv hervorgehoben, sowie die damit einhergehende Effizienzsteigerung in den Akquisitionsprozessen, und andererseits der Nutzen durch die gezielte Gestaltung des Contents und der Informationsdissemination als Vorabinformation für potenzielle Bewerber genannt.

3.2 Praxisbeispiele von Online-Intermediären

Durch Nutzung intermediärer Plattformen wie *XING* können entsprechende Botschaften, welche zum Beispiel die Nachhaltigkeit betreffen, nach außen kommuniziert werden. Zudem kann der Kontakt mit der anvisierten Zielgruppe aufgebaut und die eigene Reputation erhöht werden. Bei der im Rahmen des eRecruitings notwendigen Suche nach neuen Talenten werden die Portale *LinkedIn* und *XING* vom HRM verwendet Neben diesen dominierenden Portalen gibt es einige andere Anbieter, welche innovative Formen von Matching und Active Sourcing anbieten.

Tabelle 3 zeigt eine Kennzahlenübersicht fünf ausgewählter Plattformen aus dem E-Recruiting und E-Job Search. Diese Kennzahlen basieren auf aktuellen Daten von *SimilarWeb.com*. „Estimated Visits“ gibt eine ungefähre Anzahl an Besuchen der jeweiligen Webseite an; „Time on Site“ liefert eine Aussage zu der durchschnittlichen Zeit pro Besuch, die ein Besucher auf der Webseite verbringt; „Page Views“ entspricht der Anzahl der Seiten, die Besucher einer Webseite durchschnittlich ansehen; „Bounce Rate“ entspricht der Absprungrate, die darüber Auskunft gibt, wie groß der Anteil jener Besucher der Webseite ist, die diese sofort wieder verlassen haben.

	LinkedIn	*XING*	*BranchOut*	*Watchado*	*MercuryPuzzle*
Estimated Visits	577 Mio.	14,7 Mio.	40.000	25.000	2.000
Time on Site	6:53	5:11	1:20	4:12	6:07
Page Views	6,33	5,70	1,83	2,46	1,26
Bounce Rate (%)	32,68	41,79	55,87	59,9	73,26

Tab. 3: Kennzahlenübersicht ausgewählter E-Recruiting und E-Job Search-Plattformen (Stand: Mai 2015)

Während die Plattform *LinkedIn* der weltweite Branchenführer ist, wird im deutschsprachigen Raum besonders *XING* zur Vernetzung gewählt. Dadurch wird die Möglichkeit geschaffen ein weitreichendes berufliches Netzwerk aufzubauen. Unternehmen profitieren davon mit der angestrebten Zielgruppe zu interagieren und gezielt nach potenziellen neuen Mitarbeitern zu suchen und diese anzusprechen. So können spezielle Zielgruppen, wie beispielsweise gut ausgebildete Fachkräfte, über *XING* besonders gut erreicht werden.

Für Arbeitgeber bilden Ratingseiten eine weitere wichtige Option der Onlinepräsenz um die die Reputation im eRecruiting Bereich zu pflegen und zu kommunizieren; hierbei können aktuelle und ehemalige Arbeitnehmer ihr Unternehmen in Hinsicht auf die Arbeitgeberqualitäten bewerten. Plattformen, wie beispielsweise *kununu.com*, ermöglichen Interessierten direkte Einblicke in den Arbeitsalltag von gelisteten Unternehmen.

4 Conclusio und Ausblick

Für das Themenfeld Nachhaltigkeit im HRM-Bereich eröffnen sich durch Weiterentwicklungen in der IKT neue Chancen für Unternehmen hinsichtlich Verfügbarkeit, Erreichbarkeit und Disponibilität von Mitarbeitern. Active Sourcing wird in Abhängigkeit von der zu besetzenden Position, der Branche, den Stellenanforderungen und der Unternehmensgröße eingesetzt. eRecruiting ist bei Großunternehmen und Konzernen für Erreichbarkeit junger Zielgruppen, für eine Steigerung von Effizienz von Prozessen bei der Personalrekrutierung und zur Informationsbeschaffung von Bedeutung.

Noch vor einem Jahrzehnt steckte das Thema Nachhaltigkeitskommunikation in den Anfängen, doch mittlerweile haben sich die Erwartungshaltung und die Anforderungen im HRM-Bereich gewandelt. Eine angemessene Nachhaltigkeitskommunikation über gängige Kommunikationskanäle und –plattformen wird mittlerweile vom HRM vieler Unternehmen als notwendige begleitende Maßnahme bei der Personalrekrutierung angesehen, um sich bei der anvisierten Zielgruppe als attraktiver Arbeitgeber zu positionieren und um im Wettbewerb um qualifiziertes Personal Vorteile zu erlangen.

Literaturverzeichnis

[Ba12] Bauer, T.N., Erdogan, B., Taylor, S.: Creating and maintaining environmentally sustainable organizations: Recruitment and onboarding, 2012.

[Bh08] Bhattacharya, C.B., Sen, S., Korschun, D.: Using corporate social responsibility to win the war for talent. MIT Sloan Management Review, 49, S. 37–44, 2008.

[Bo05] Boudreau, J. W., Ramstad, P. M.: Talentship, talent segmentation, and sustainability: A new HR decision science paradigm for a new strategy definition. Human Resource Management, 44, S. 129–136, 2005.

[Ch05] Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A., Jones, D. A.: Applicant attraction to organizations and job choice: a meta-analytic review of the correlates of recruiting outcomes. Journal of applied psychology, 90(5), S. 928, 2005.

[Cl13] Clement, R., Schreiber, D.: Internet-Ökonomie: Grundlagen und Fallbeispiele der vernetzten Wirtschaft. Springer-Verlag, 2013.

[Co12] Cohen, E., Taylor, S., Muller-Camen, M.: HRM's role in corporate social and environmental sustainability. Research Report, SHRM, 2012.

[Di07] Dineen B. R., Ling J., Ash S. R., Del Vecchio D.: Aesthetic properties and message customization: Navigating the dark side of web recruitment. Journal of Applied Psychology 92, 2007.

[Ec13] Eccles, R. G., Serafeim, G.: The performance frontier: Innovating for a sustainable strategy. Harvard Business Review, 91, S. 50–60, 2013.

[Ec14] Eckhardt, A., Laumer, S., Maier, C., Weitzel, T.: The transformation of people, processes, and IT in e-recruiting: Insights from an eight-year case study of a German media corporation. Employee Relations, 36(4), S. 415-431, 2014.

[Eh14] Ehnert, I., Harry, W., Brewster, C.: Sustainable HRM in Europe: Diverse contexts and multiple bottom lines. In I. Ehnert, W. Harry, K. J. Zink (Eds.), Sustainability and human resource management: Developing sustainable business organizations. S. 339–357. Heidelberg: Springer, 2014.

[Gh14] Ghazzawi, K., Accoumeh, A.: Critical Success Factors of the E-Recruitment System. Journal of Human Resources, 2(2), S. 159-170, 2014.

[Gl95] Gladwin, T. N., Kennelly, J. J., Krause, T. S.: Shifting paradigms for sustainable development: Implications for management theory and research. Academy of Management Journal, 20, S. 874–907, 1995.

[Gr15] GRI: About GRI, https:// www.globalreporting.org/Pages/default.aspx, 09.05.2015.

[Gu14] Guerci, M., Pedrini, M.: The consensus between Italian HR and sustainability managers on HR management for sustainability-driven change – towards a 'strong' HR management system. The International Journal of Human Resource Management, 25, S. 1787–1814, 2014.

[Ha14] Hahn, R., Lülfs, R.: Legitimizing negative aspects in GRI-oriented sustainability reporting: A qualitative analysis of corporate disclosure strategies. Journal of Business Ethics, 123, 401–420, 2014.

[Hi02] Highhouse, S., Hoffman, J. R., Greve, E. M., Collins, A. E.: Persuasive Impact of Organizational Value Statements in a Recruitment Context1. Journal of Applied Social Psychology, 32(8), 1737-1755, 2002.

[Ho14] Hofmann, H., Busse, C., Bode, C., Henke, M.: Sustainability-related supply chain risks: Conceptualization and management. Business Strategy and the Environment, 23, 160–172, 2014.

[Is15] Isenmann, R.: Zur Zukunft der Nachhaltigkeitsberichterstattung. Neue Wege der Dokumentation und Kommunikation. Business Reporting. Zeitschrift für Berichterstattung und Dokumentation 1(4): 112-116, 2015.

[Ja15] Jaksch, D.: Recruiting 2.0 – onlinebasiertes Recruiting von internationalen Unternehmen am Standort Wien. Universität Wien, 2015.

[Kp13] KPMG: The KPMG survey of corporate responsibility reporting 2013. Amsterdam: KPMG. International Cooperatives, 2013.

[Li15] LinkedIn, https://www.linkedin.com/, 01.06.2015

[Me13] Meurer S., Jäger W., Böhm S.: Mobile Recruiting 2013, Eine empirische Untersuchung zur Bewerberansprache über mobile Endgeräte. Wiesbaden, 2013.

[Me15] MercuryPuzzle, https://mercurypuzzle.com/, 10.05.2015

[M11] Mladenow, A., Fröschl, K. A.: Kooperative Forschung. Frankfurt am Main: Lang, 2011.

[Na15] Nachhaltigkeit.at: https://www.nachhaltigkeit.at/thema/corporate-social-responibility/csr-auf-europaeischer-ebene/april-2014-eu-richtlinie-zur-verpflichtenden-nachhaltigkeitsberichterstattung, 10.05.2015

[No08] Nohl, A.M.: Interview und dokumentarische Methode: Anleitungen für die Forschungspraxis. Springer, 2008.

[Re15] Respact, https://www.respact.at/site/themen/nachhaltigkeitsberichters/article/6577.html 10.05.2015.

[Sc09a] Schreurs B., Derous E., Van Hooft E. A. J., Proost K., De Witte K.: Predicting applicants' job pursuit behavior from their selection expectations: the mediating role of theory of planned behavior. In: Journal of Organizational Behavior. Vol. 30. No. 6. S. 761–783, 2009.

[Sc09b] Schuhmacher, F., Geschwill, R.: Employer Branding (pp. 37-51). Gabler, 2009.

[Se09] Senna, J., Shani, A. B.: Utilizing technology to support sustainability. In P. Docherty,M. Kira, & A. B. Shani (Eds.), Creating sustainable work systems: Developing social sustainability (pp. 84–100). London: Routledge, 2009.

[St91] Strauss, C.: Informatik-Sicherheitsmanagement. BG Teubner, Stuttgart, 1991.

[St01] Strauss, C.: Quantitative Personaleinsatzplanung im Airline Business. Peter Lang Verlag, Frankfurt am Main, 2001.

[Th08] Thompson L. F., Braddy P. W., Wuensch K. L.: E-recruitment and the benefits of organizational web appeal. Computers in Human Behaviour 24/2008, 2008.

[Wh15] Turban, D. B., Cable, D. M. Firm reputation and applicant pool characteristics. Journal of Organizational Behavior, 24(6), 733-751, 2003.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Die Automatisierung von Compliance in Cloud-Umgebungen

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Die Automatisierung von Compliance in Cloud-Umgebungen

Prof. Dr. Ing. Rainer Böhme[1] und Ulrich Waldmann[2]

Abstract: Der positive Trend zum Cloud Computing ist ungebrochen. Dennoch zögern viele potenzielle Anwender, den Schritt in die Cloud zu wagen und dadurch die Vorteile von Cloud-Services zu nutzen. Viele Unternehmen haben Bedenken bezüglich der Gewährleistung von Datenschutz und Datensicherheit sowie der Einhaltung von gesetzlichen Regelungen (Compliance). Es besteht somit ein großer gesellschaftlicher und wirtschaftlicher Bedarf nach Lösungen, die es Anwendern ermöglichen, die Vertrauenswürdigkeit einer Cloud-Anwendung beurteilen zu können. Die Messung der Vertrauenswürdigkeit muss dabei so weit wie möglich automatisiert erfolgen.

Der Workshop „Automatisierung von Compliance in Cloud-Umgebungen“ und seine Beiträge befassen sich mit der automatisierten Überprüfbarkeit von Datenschutz- und Datensicherheitseigenschaften von Cloud-Diensten. Einen Schwerpunkt bilden dabei Methoden, die auf Kennzahlen basieren oder die Berechnung von Kennzahlen ermöglichen. Die Beiträge sind sowohl technischer als auch wirtschaftlicher und rechtlicher Natur.

Für die Vorschläge der geladenen Vorträge und die Begutachtung der eingereichten Beiträge in diesem Tagungsband danken wir den Mitgliedern des Programmkomitees:

- Prof. Rainer Böhme (Westfälische Wilhelms-Universität Münster),
- Michael Herfert (Fraunhofer SIT),
- Prof. Martin Kütz (Hochschule Anhalt),
- Prof. Stefan Sackmann (Martin-Luther-Universität Halle-Wittenberg),
- Dr. Aleksandra Sowa (Deutsche Telekom),
- Prof. Gottfried Vossen (Westfälische Wilhelms-Universität Münster),
- Dr. Karl-Friedrich Thier (T-Systems International),
- Prof. Michael Waidner (Technische Universität Darmstadt und Fraunhofer SIT),
- Prof. Thomas Wilmer (Hochschule Darmstadt).

Die Organisatoren des Workshops danken dem BMBF, welches einige der diesem Workshop zugrunde liegenden Forschungsarbeiten im Rahmen des Projekts VeriMetrix im Programm IKT 2020 – Forschung für Innovationen fördert (Förderkennzeichen: 16KIS0053K).

[1] Westfälische Wilhelms-Universität Münster, Institut für Wirtschaftsinformatik, Leonardo-Campus 3, 48149 Münster, rainer.boehme@uni-muenster.de

[2] Fraunhofer-Institut für Sichere Informationstechnologie SIT, Rheinstraße 75, 64295 Darmstadt, ulrich.waldmann@sit.fraunhofer.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Evidente Ausfallrisiken im Cloud-Markt – Eine quantitative Analyse der Finanzberichte von Anbietern

Andreas Jede[1], Frank Teuteberg

Abstract: Der Cloud-Markt verzeichnet nach wie vor dynamische Wachstumsraten und verspricht den Nutzern günstige Unterstützung bei diversen Geschäftsprozessen. Dabei wird das Cloud-Angebot durch die steigende Zahl der Anbieter sowie durch die steigenden Service-Verflechtungen zunehmend intransparenter. Weitergehend ist die Anbieterseite von einer hohen Vielfalt gekennzeichnet, da sich sowohl etablierte als auch junge oder weitestgehend unbekannte IT-Unternehmen, die i.d.R. eine geringere Finanzkraft mitbringen, um den Cloud-Markt bemühen. Die Forschung hat zwar wertvolle und differenzierte Rahmenwerke hervorgebracht, die die IT-Abteilungen der nutzenden Unternehmen im Cloud-Selektionsprozess begleiten. Die Dimension „Anbieterausfallrisiko" wurde jedoch bislang eher nebensächlich diskutiert, was eine Forschungslücke darstellt. Gerade vor dem Hintergrund des vielschichtigen Begriffs „Vertrauenswürdigkeit" gilt es finanziell gesunde Anbieter von Risikoanbietern zu separieren. Diese Arbeit analysiert systematisch die Finanzberichte von 38 Cloud-Anbietern. Die Ergebnisse indizieren die dringende Notwendigkeit zur Schaffung von aktiven Steuerungsinstrumenten für Anbieterausfallrisiken. Die Theorie profitiert von diesem Beitrag durch die Ergänzung bestehender Rahmenwerke um die wichtige Zusatzdimension „Anbieterausfallrisiko" samt eines Berechnungsschemas.

Keywords: Cloud Computing, Compliance, Ausfallrisiken, Finanzanalyse, Inhaltsanalyse

1 Einleitung

Ressourcenschonung, höhere Wettbewerbsfähigkeit und finanzielle Vorteile sind gewichtige Aspekte, die dem Paradigma „Cloud Computing" (CC) in den vergangenen Jahren enorme Wachstumsraten bescherten und ein Ende des „Cloud-Booms" ist nicht absehbar [Ma11, Ri14]. Dabei übermitteln CC-Nutzer (sensible) Daten an die CC-Anbieter in der Annahme, dass Datensicherheit und Datenschutz gewährleistet sind. Aufgrund der unilateralen Abhängigkeit seitens der CC-Nutzer von den CC-Anbietern spielt der Faktor „Vertrauen" eine wesentliche Rolle [Wa13]. In der Literatur hat sich keine allgemeine Definition für den Vertrauensbegriff durchgesetzt [Ro98]. Während sich beispielsweise Psychologen und Soziologen Bereichen wie Erwartungen, Beziehungen und Charakteristiken widmen, fokussieren sich Ökonomen eher auf quantitative oder institutionelle Untersuchungen, was auf die Komplexität und Vielschichtigkeit des Begriffs schließen lässt. Neuere Arbeiten der Wirtschaftsinformatik subsummieren unter den Vertrauensbegriff sowohl harte Faktoren wie z.B. Funktionalität und Stabilität als auch weichere Faktoren wie Glaubwürdigkeit und Verlässlichkeit [Wa13]. Für die Be-

[1] Universität Osnabrück, Fachbereich Unternehmensrechnung und Wirtschaftsinformatik, Katharinenstraße 1, 49069 Osnabrück, andreas-jede@gmx.de

wertung der Vertrauenswürdigkeit von CC-Anbietern hat die wissenschaftliche Forschung zahlreiche Arbeiten und Rahmenwerke hervorgebracht, die sowohl die Chancen als auch die mit der Cloud-Nutzung einhergehenden Risiken (z.B. Datensicherheit, Lock-In-Effekte, rechtliche Rahmenbedingungen, proprietäre Software, häufige Service-Updates, Schnittstellenprobleme) thematisieren [z. B. MC13, Du13, KK14]. Im Rahmen der bisherigen wissenschaftlichen Literatur wird ein gravierendes Risiko jedoch weitestgehend außer Acht gelassen, welches einen interdisziplinären Ansatz zwingend erfordert: die schwache Finanzlage und Validität zahlreicher, vor allem kleiner CC-Anbieter. Gerade die extrem hohe Dynamik des CC-Marktes bringt viele neue Anbieter mit einer breiten CC-Produktpalette auf den Plan, die einem höheren Ausfallrisiko unterliegen als traditionelle IT-Anbieter mit einem diversifizierten Produktportfolio. Ferner sind zahlreiche Anbieter komplett fremd- oder unterfinanziert, was im Allgemeinen einen risikofreudigeren Geschäftssinn indiziert und somit höhere Insolvenzrisiken birgt [At12]. So stellten in jüngster Vergangenheit mehrere CC-Anbieter (z.B. Nirvanix, Nimbula) ihre Dienstleistung konkursbedingt ohne zeitgerechte Vorwarnungen an ihre Kunden ein. Je nach Bedeutung und Art der Services können die Nutzer erhebliche Auswirkungen erleiden, bis hin zur gezwungenen Unterbrechung ihrer Geschäftsfähigkeit. Erschwerend kommt hinzu, dass sich hinter einem CC-Service in den seltensten Fällen ein einziger Anbieter verbirgt [Le10]. Auch wenn der Nutzer den Vertrag mit einem Anbieter abschließt, besteht der Service in der Regel aus einem komplexen CC-Netzwerk mit diversen vorgelagerten Anbietern, die auf einen Teilbereich des CC-Services spezialisiert sind (z. B. Software, Plattform). Bei einem dauerhaften Ausfall eines relevanten Teilbereichs kann der gesamte Service zum Erliegen kommen. Daher sind bei einem Anbieterausfall nicht nur Nutzer sondern auch andere abhängige Anbieter im CC-Netzwerk von den Auswirkungen direkt betroffen. Um als Nutzer die Risiken einer Abhängigkeit von insolvenzgefährdeten CC-Anbietern zu minimieren, sollte die unternehmerische Vertrauenswürdigkeit der CC-Anbieter vorab bewertet werden. Aus diesem Grund untersucht diese Arbeit die finanzielle Lage ausgewählter CC-Anbieter. Hierdurch sollen die Ausfallrisiken eingeordnet und die bisherigen Arbeiten und Rahmenwerke um eine wichtige Risikodimension ergänzt werden. Der zugrundeliegende wissenschaftliche Ansatz basiert auf der Methode der systematischen Dokumentenanalyse [Fe06]. Die Forschungsfrage lautet dabei: *Lassen sich evidente Ausfallrisiken aus den Finanzberichten der CC-Anbieter ableiten und wie können diese Risiken im CC-Selektionsprozess geeignet Beachtung finden?* Die Arbeit ist wie folgt aufgebaut: Nach der Einleitung folgt im zweiten Abschnitt eine Erläuterung der begrifflichen Grundlagen. Anschließend beschreibt der dritte Abschnitt das methodische Vorgehen der Arbeit. Im vierten Abschnitt werden die Ergebnisse der Analysen vorgestellt, die im fünften Abschnitt zu Implikationen für Wissenschaft und Praxis führen. Die Arbeit schließt mit einem Fazit im letzten Abschnitt.

2 Begriffliche Grundlagen

CC wird vom „National Institute of Standards and Technology" wie folgt definiert [MG11]: „a model for enabling ubiquitous, convenient, on-demand network access to a

shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction". Dabei stellt CC keine neuartige Technologie dar, sondern ein neues Paradigma für IT-Prozesse, indem es bekannte Einzeltechnologien konsequent verknüpft [Le10]. Ferner wird in der Literatur zwischen drei „as a Service"-Modellen unterschieden [MG11, Le10]: Infrastructure, Platform und Software (IaaS, PaaS, SaaS). Desweiteren existieren vier CC-Ausprägungen [Ma11]: Der Ursprung geht auf das „Public CC" zurück, bei dem ein externer Anbieter seine Services über das Internet zur Verfügung stellt. Beim „Private CC" hingegen werden Services unternehmensintern angeboten, was implizit einen notwendigen, eigenen CC-Betrieb bedeutet. Die dritte Version „Community CC" ist in Betracht zu ziehen, falls eine Gruppe von Unternehmen, die gemeinsame Interessen verfolgt und vergleichbare Sicherheitsstandards pflegt, den CC-Betrieb eigenständig kontrollieren will. Die vierte Version „Hybrid CC" stellt eine Kombination aus „Public CC" und „Private CC" dar. In der Regel werden dabei sensible Informationen intern abgegrenzt und unkritische Services und Daten in die Obhut eines externen Anbieters übertragen. Im Rahmen dieser Arbeit sind insbesondere Anbieter des „Public CC" relevant, da die Abhängigkeit bei externen Anbietern am höchsten ist. Weiterhin steht hinter einem Cloud-Service in der Regel ein Netzwerk, welches mehrere Anbieter enthalten kann [Le10]. Neben den drei oben beschriebenen „as a Service" Anbietern kann das Netzwerk auch Integratoren, Aggregatoren oder Mediatoren enthalten, die alle eine gewisse Wertschöpfung innerhalb des Netzwerks erbringen und somit voneinander abhängig sind [Wa14].

3 Methodisches Vorgehen

3.1 Dokumentenbasis

Der vorliegende Beitrag adaptiert die Methode der systematischen Literaturanalyse, die sich in der wissenschaftlichen Forschung zweckdienlich zur Identifikation eines aktuellen Problems erwiesen hat [Fe06]. Die Dokumentenbasis stellen hierbei die Finanzberichte von CC-Unternehmen dar. Die Durchführung der Analyse erfolgt in enger Anlehnung an das Fünf-Stufen-Modell von Fettke [Fe06]: 1) Problemformulierung, 2) Dokumentensuche, 3) -auswertung, 4) -analyse sowie 5) Interpretation. Hierbei ist hervorzuheben, dass ausgehend von der Forschungsfrage in der Einleitung (Stufe 1) die Suche (Stufe 2) nach relevanten CC-Anbietern basierend auf drei Ranglisten (www.**talkincloud**.com/tc100; www.**cloudtimes**.org/top100/; http://www.**forbes**.com/sites/louiscolumbus/2015/01/29/the-best-cloud-computing-companies-and-ceos-to-work-for-in-2015/) erfolgte. Diese drei Listen führten zu 157 verschiedenen Unternehmen, wobei folgende drei Anbieterkategorien exkludiert wurden: (i) Das Unternehmen erzielt *nicht den überwiegenden Anteil seines Umsatzes mit CC* (> 75%), sondern durch andere Einnahmequellen wie z.B. traditionelle Dienstleitungen oder Werbung (z.B. SAP, Oracle). (ii) Das Unternehmen bietet *vor allem CC-Security-Services* an (z.B. FireEye, Cyren, Wix.com). Denn der Ausfall dieser Dienstleister geht in der Regel nicht mit einem grö-

ßeren Datenverlust einher und diese Anbieter sind meist leichter zu wechseln. (iii) Das Unternehmen ist *nicht-börsennotiert*. Da nur börsennotierte Unternehmen verpflichtet sind, ihre Abschlüsse zu veröffentlichen, konnte in der Gegenprobe kein nicht-börsennotierter Anbieter gefunden werden, der die Abschlüsse freiwillig veröffentlicht. Diese drei Selektionsmerkmale führten schließlich zu 31 relevanten CC-Anbietern. Ergänzend dazu wurden im zweiten Schritt alle Technologiewerte geprüft, die an den Börsen in New York, London und Frankfurt gelistet sind und die obigen Kriterien erfüllen, was die Anzahl der relevanten CC-Unternehmen auf 38 erhöhte[2].

3.2 Dokumentenauswertung

Für die Auswertung und Finanzanalyse (Stufen 3 und 4 [Fe06]) wurden die aktuellen Jahresabschlüsse der 38 Unternehmen berücksichtigt. Bei der Berechnung des Ausfallrisikos lehnt sich diese Arbeit an das klassische Z-Faktoren-Modell von Altman [Al68] an. Auch wenn zahlreiche Arbeiten das Modell infrage stellen [z. B. Be05, S.75], genießt es nach wie vor hohes Ansehen in der wissenschaftlichen Forschung und bildet in der Praxis oftmals die Grundlage namhafter Ratingagenturen [Ol12, WC10]. In seinem Modell parametrisiert Altman mithilfe der multivariaten, linearen Diskriminanzanalyse einen Z-Faktor und vergleicht die zu beurteilenden Unternehmen mit einer „voraussichtlich solventen“ Alternativgruppe (Bestimmung des Signifikanzniveaus). Er erzielt mit seinem Modell eine Prognosegenauigkeit für Insolvenzen von 95% bei einem Horizont von einem Jahr und 72% bei zwei Jahren. Zahlreiche Autoren nutzten Altmans Modelle als Basis und führten spezifische Anpassungen durch (z. B. für bestimmte Länder oder Branchen), um die Prognosequalität weiter zu erhöhen. Da im vorliegendem Fall 97% der selektierten CC-Anbieter amerikanischer Herkunft sind und der Fokus dieser Arbeit eher auf dem Schaffen eines grundsätzlichen Bewusstseins und weniger auf präzisen Prognosen bezüglich spezifischer Anbieter liegt, bedient sich diese Arbeit der originären Z-Faktoren. Dem Verfahren von Altman folgend wurde eine Alternativgruppe selektiert, die ebenfalls aus 38 börsennotierten Unternehmen besteht. Für die Suche nach alternativen IT-Anbietern wurden IT-Unternehmen aus dem Index „S&P 500“ herangezogen und nach Herkunftsland „USA“ gefiltert. Hierbei wurden Unternehmen mit größerem CC-Angebot selbstverständlich exkludiert. Danach wurden manuell diejenigen IT-Unternehmen selektiert, die gemäß ihrem Anlagevermögen den einzelnen CC-Unternehmen der Fokusgruppe am ähnlichsten sind. Um das Konstrukt auf seine Validität zu testen, wurden die beiden Gruppen mit je 38 Unternehmen schließlich nach dem Zufallsprinzip in zwei Sub-gruppen geteilt: Die „Schätzproben“ bestehen dabei aus 28 CC-Unternehmen und 28 Alternativunternehmen und die „Prognoseproben“ aus 10 CC-Unternehmen und 10 Alternativunternehmen. Das beschriebene Vorgehen folgt den Ansätzen von Altman [Al68, Al00], der die Wirksamkeit seiner Z-Faktoren mit lediglich 33 Unternehmen in der Fokusgruppe bewiesen hat.

[2] Eine Liste der 38 CC-Unternehmen ist verfügbar unter:
https://ssl.tsdprivatserver.de/share/1427016413/38 CC-Anbieter.pdf

3.3 Quantitative Inhaltsanalyse

Ferner wurde eine quantitative Inhaltsanalyse durchgeführt, die auf die Identifikation detaillierter Zusammenhänge zwischen den Termini sowie auf die Identifikation verborgener Merkmale abzielt [Su01]. Dazu wurden die Berichte der 38 CC-Anbieter herangezogen. Für die Inhaltsanalyse wurde die Software „WordStat“ von Provalis genutzt. Die Berichte werden dabei in die Software kopiert und in vier Schritten bearbeitet. Während der erste Schritt die manuelle Normierung der Schreibweise und der Silbentrennung sowie die Entfernung aller Klammern erforderte, beinhaltete der zweite Schritt die systemgestützte Lemmatisierung, die Stammformreduktion, sowie das Exkludieren unnötiger Füllwörter. Die gewonnene Datenbasis diente anschließend der Analyse und der Interpretation. Zum Vergleich wurden auch die 38 Berichte der Alternativgruppe analysiert. Alle Berichte sind in englischer Sprache verfasst.

3.4 Themenverwandte Veröffentlichungen

In einer weiteren systematischen Literatursuche wurden fünf Arbeiten identifiziert, die ein Rahmenwerk für Risikobetrachtungen im CC-Paradigma abhandeln. Diese Arbeiten sind in der Tabelle 1 jeweils mit „Forschungsfrage“ und „Ergebnissen“ aufgeführt.

Ref.	Forschungsfrage	Ergebnisse
[Ar10]	Wie ist CC definiert und was sind Chancen und Hindernisse?	Risikorahmenwerk besteht aus 10 Hindernissen: Service-Verfügbarkeit, Lock-In-Effekte, Datensicherheit, Datentransfer, Performanceunsicherheit, Datenspeicherung, Bugs, Skalierung, Reputation anderer Nutzer, Lizenzsystem.
[Gr11]	Was sind die Schwachstellen bei der CC-Nutzung?	Basierend auf einer anerkannten Risikotaxonomie werden folgende Schwachstellen analysiert: Kerntechnologie (z.B. http-Protokolle, kryptographische Eigenschaften); CC-Charakteristiken (z.B. Daten-Recovery, Bezahlsystem); Sicherheitskontrollen; Gefahren bei speziellen CC-Angeboten (z.B. anfälliges „Cross-site scripting“ oder „SQL-injection“).
[MT12]	Wie kann ein Entscheidungssystem in CC-Umgebungen ausgestaltet sein?	Mathematisches Modell, welches die CC-Risiken in drei Kategorien gruppiert: Integrität, Service-Verfügbarkeit und Vertraulichkeit.
[Du13]	Welche CC-Risiken entstehen für die nutzende, gesamte Organisation?	Aus der Literatur hergeleitetes Rahmenwerk mit fünf übergeordneten Kategorien (IT-Organisation; CC-Betrieb; Technologie; Rechtliche Risiken), die jeweils aus drei bis fünf Unterkategorien bestehen.
[KK14]	Wie können Risiken in Netzwerken von CC-Anbietern identifiziert werden?	Zunächst Erstellung eines Referenzmodells mit Akteuren in CC Netzwerken. Anschließend Verknüpfung der Akteure mit den jeweiligen Risiken (Datenrisiken, Preisrisiken sowie Verfügbarkeitsrisiken).

Tabelle 1: Themenverwandte Veröffentlichungen

Dieser Beitrag grenzt sich von den bisherigen Arbeiten wie folgt deutlich ab: die zu hinterfragende Finanzstärke sowie die Gefahr einer Insolvenz von CC-Anbietern sind bis

dato eher rudimentär und in allgemeinen, kurzen Aussagen ohne fundierte Analysen angeführt worden [z. B. Ma11, Ar10]. Zwar diskutiert die Literatur die Service-Verfügbarkeit als eine Risikogröße. Damit sind aber technische Faktoren wie z.B. „Response time“ oder „99,X% Leistungserbringung“ gemeint. Diese Arbeit eruiert die Notwenigkeit einer Erweiterung der Literatur durch die Schaffung der zusätzlichen Dimension „Anbieterinsolvenz“ bei der Selektion geeigneter CC-Anbieter. Damit soll die Vertrauenswürdigkeit der Anbieter aus finanzieller Perspektive indiziert werden.

4 Auswertung und Analyse

4.1 Ergebnisse der Finanzanalyse

Die Diskriminanzfunktion für börsennotierte, amerikanische Unternehmen, die nicht im Finanzsektor (z. B. Banken) tätig sind, lautet gemäß Altman [Al68]: $Z´= 1,2X_1+1,4X_2+3,3X_3+0,6X_4+0,999X_5$. Die Variablenausprägungen sind der Tabelle 2 zu entnehmen. Unternehmen, die einen Z-Faktor von weniger als 1,81 besitzen, werden dem Modell nach als stark insolvenzgefährdet angesehen. Bei einem Wert von größer als 2,67 ist das Unternehmen hingegen ungefährdet. In einer von vielen späteren Arbeiten entwickelte Altman [Al00] einen Z-Faktor speziell für Unternehmen in aufstrebenden und neuen Märkten, was gewisser Maßen auch für CC-Unternehmen gilt. Diese Funktion lautet: $Z´´= 6,56X_1+3,26X_2+6,72X_3+1,05X_4$. Die kritische untere Grenze liegt hier bei 1,10. Mittels beider Funktionen wurden die Finanzberichte manuell ausgewertet.

Variable	Berechnung
X_1	(Umlaufvermögen – kurzfristige Verbindlichkeiten) / Bilanzsumme
X_2	Einbehaltene Gewinne / Bilanzsumme
X_3	Ergebnis vor Zinsen und Steuern / Bilanzsumme
X_4	Marktwert des Eigenkapitals / Summe der Verbindlichkeiten
X_5	Umsatz / Bilanzsumme

Tabelle 2: Variablen des Z-Faktor-Modells

Wie zuvor beschrieben wurden die 38 selektierten CC-Unternehmen (Fokusgruppe) per Zufallsprinzip in eine Prognoseprobe und eine Schätzprobe aufgeteilt und mit den jeweiligen Proben der Alternativgruppe verglichen. Die Ergebnisse sind in der Tabelle 3 aufgeführt. So zeigen die Durchschnittswerte der CC-Unternehmen in beiden Proben erhebliche Insolvenzrisiken, da die unteren Grenzwerte beider Z-Faktoren deutlich unterschritten sind. Zwar gibt es auch bei den CC-Anbietern „gesunde“ Unternehmen in beiden Proben (siehe Maximum), diese sind jedoch in der Minderheit. Vor allem die Variablen X_2 und X_3 haben einen starken negativen Einfluss auf die Z-Faktoren. Diese Variablen messen die vergangenen und aktuellen Gewinne der CC-Unternehmen, die bei den CC-Unternehmen überwiegend negativ ausfallen. Damit können sich diese Unternehmen nicht eigenständig refinanzieren und sind stark vom „Wohlwollen“ ihrer Investoren abhängig. Dabei enthält die Fokusgruppe nicht nur Start-Ups sondern auch gesetzte CC-

Unternehmen. Im Durchschnitt sind die 38 CC-Unternehmen seit 15 Jahren aktiv, womit das Argument der „Anschubinvestitionen“ entkräftet werden kann. Die Werte der Alternativgruppe weisen auf ein entgegengesetztes Bild hin. In beiden Proben sind zwar ebenfalls insolvenzrisikobehaftete Unternehmen enthalten, diese sind jedoch in der Minderheit. Die Durchschnittswerte beider Proben der Alternativgruppe sind nicht kritisch.

Art der Probe	Statistische Größe	X_1	X_2	X_3	X_4	X_5	Z´	Z´´
Schätzprobe der Fokusgruppe (n=28)	Durchschnitt	0,09	-0,43	-0,04	1,67	0,54	**0,89**	**0,61**
	Minimum	0,01	-4,03	-0,32	-0,37	0,22	**-6,69**	**-15,60**
	Maximum	0,24	0,28	0,11	5,53	1,39	**5,73**	**8,99**
	Standardabweichung	0,06	0,79	0,12	1,50	0,24	**2,70**	**5,31**
Schätzprobe der Alternativgruppe (n=28)	Durchschnitt	0,10	0,20	0,08	1,10	0,60	**1,92**	**2,98**
	Minimum	-0,03	-1,49	-0,28	-0,19	0,26	**-2,90**	**-7,13**
	Maximum	0,25	1,80	0,39	4,11	1,53	**8,09**	**14,42**
	Standardabweichung	0,07	0,62	0,12	0,80	0,27	**2,10**	**4,13**
Prognoseprobe der Fokusgruppe (n=10)	Durchschnitt	0,05	-0,67	-0,07	1,11	0,51	**0,06**	**-1,17**
	Minimum	-0,06	-2,80	-0,32	0,12	0,19	**-4,80**	**-11,58**
	Maximum	0,13	0,31	0,12	2,79	0,93	**3,60**	**5,64**
	Standardabweichung	0,06	0,82	0,13	0,79	0,17	**2,28**	**4,73**
Prognoseprobe der Alternativgruppe (n=10)	Durchschnitt	0,13	0,37	0,08	3,97	1,00	**4,31**	**6,74**
	Minimum	0,04	-0,37	-0,07	0,53	0,27	**-0,09**	**-0,83**
	Maximum	0,51	1,28	0,46	6,95	3,02	**11,09**	**17,87**
	Standardabweichung	0,13	0,45	0,13	2,21	0,79	**3,36**	**5,57**

Tabelle 3: Statistische Größen zur Berechnung der Insolvenzrisiken

4.2 Ergebnisse der statistischen Signifikanz

Zur Bestimmung der Wahrscheinlichkeitsverteilung wurden die Tests von Kolmogorov-Smirnov und Shapiro-Wilk angewandt. Beide Tests indizierten eine Normalverteilung der Variablen in den jeweiligen Gruppen, womit die Voraussetzungen für einen Zweistichproben-t-Test erfüllt sind. Dieser Test prüft anhand der Durchschnittswerte zweier unabhängiger Stichproben, in welcher Relation die Durchschnittswerte zweier Grundgesamtheiten zueinander stehen. Da die Grundgesamtheiten sowohl in den beiden Schätzproben als auch in den Prognoseproben nicht der gleichen Varianz entstammen, wurde eine spezielle Variante von Zweistichproben-Tests namens „Welch“ herangezogen, die die Gleichheit der Varianzen nicht voraussetzt. Anhand des Welch-Tests wurden schließlich die p-Werte ermittelt, um das Signifikanzniveau zu prüfen. Die Nullhypothese lautet, dass die Durchschnittswerte der Proben jeweils derselben Grundgesamtheit entspringen. In der Tabelle 4 sind die p-Werte dargestellt, wobei folgende Symbolik gilt: ** für $p<0,01$,* für $p< 0,05$ und *n.s.* für nicht signifikant. Als äußerst signifikant erweisen sich die Variablen X_2, X_3 und X_4. Bei den Variablen X_1 und X_5 kann die Nullhypothese nicht verworfen werden. Insgesamt bestätigen auch die statistischen Ergebnisse die Hypothese, dass zwischen CC-Anbietern und Anbietern der Alternativgruppe teils signifikante Unterschiede bezüglich der Insolvenzrisiken existieren.

Art der Probe	X_1	X_2	X_3	X_4	X_5
Schätzprobe	0,095 *n.s.*	0,001**	0,000**	0,045*	0,137 *n.s.*
Prognoseprobe	0,044*	0,002**	0,013*	0,002**	0,048*

Tabelle 4: Signifikanz der Variablen

4.3 Ergebnisse der quantitativen Inhaltsanalyse

Die quantitative Inhaltsanalyse ermöglicht die Bestimmung von Worthäufigkeiten und Kookkurrenzen der Datenbasis. Die Worthäufigkeiten lassen sich aus der absoluten Summe ermitteln, wobei im Rahmen dieser Analyse die 100 meistgenannten Wörter berücksichtigt wurden. Darauf basierend wurden die Kookkurrenzen errechnet, welche in der Linguistik als ein Maß zum gemeinsamen Auftreten zweier Wörter (1:1) genutzt werden [Ta05, S.65-72]. In dieser Arbeit wurden die Kookkurrenzen auf Paragraphenbasis ermittelt. Zur Berechnung dieser Interrelationen wurde der Jaccard-Index verwendet [Ta05, S.65-72]. Aus diesen Daten können Dendrogramme abgeleitet werden, die die vorgegebenen Wörter in Verbindung mit Variablen abbilden (1:n). Die Ergebnisse können durch das Verfahren der multidimensionalen Skalierung (MDS) zu einer 2-D Karte visualisiert werden [Pr10]. Die Größe der Kreise korrespondiert mit der Worthäufigkeit. Die Distanz zwischen den Kreisen entspricht der Häufigkeit des gemeinsamen Auftretens von Wörtern. Die Farben markieren übergeordnete Gruppen, die besonders starke Abhängigkeiten repräsentieren. Die Ergebnisse sind in der Abbildung 1 zu sehen.

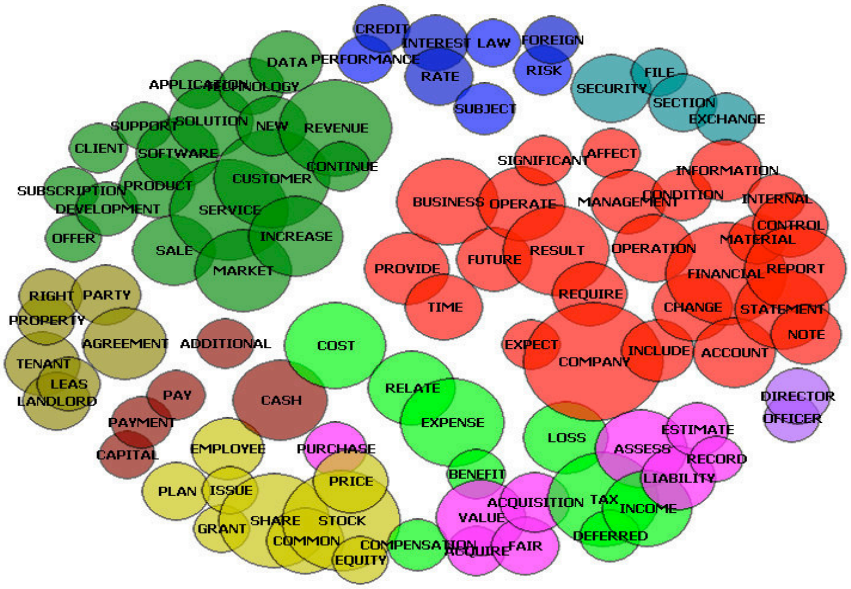


Abbildung 1: 2-D Karte der quantitativen Inhaltsanalyse von Berichten der CC-Unternehmen

Die Abbildung 1 zeigt im oberen Bereich, dass das Thema Risiko derzeit vor allem im Zusammenhang mit *Kreditrisiken* und *Zinsrisiken* verbinden, was die zuvor präsentierten Ergebnisse bestätigt. Im Falle einer Insolvenz ist völlig unklar, wie sich die Daten vor unberechtigtem Zugriff schützen lassen. In diesem Fall könnte auch der Zugang zu den Daten erschwert oder verhindert werden, was zu ungeklärten gesetzlichen Fragestellungen führt. Weiterhin werden sowohl *Performancerisiken* als auch *rechtliche Risiken* und *Sicherheitsrisiken* diskutiert. Diese drei Risiken werden in der Literatur ausführlich diskutiert (siehe Tabelle 1). Um diesen Befund abzusichern, wurden auch die 38 Anbieter der Alternativgruppe nach demselben Verfahren analysiert. Diese Ergebnisse implizieren zumindest keinen unmittelbaren Zusammenhang zu offenkundigen, finanziellen Risiken. Das fortgeschrittene Stadium der Alternativgruppe wird weiterhin dadurch deutlich, dass beim Thema Risiko vor allem der aktive Umgang mit selbigem diskutiert wird, da Termini wie *Risikoreporting*, *Risikokontrolle* sowie *Risikomanagement* vermehrt auftreten.

5 Implikationen für Theorie und Praxis

Diese Arbeit bietet zahlreiche Implikationen für die weitere wissenschaftliche Forschung. Zunächst wird deutlich, dass der ***interdisziplinäre Austausch*** zwischen Ökonomie und Informatik zwingend erforderlich ist, um Themengebiete wie Vertrauenswürdigkeit und Risikobewertung von CC-Anbietern adäquat sowohl aus technischer als auch aus ökonomischer Perspektive zu untersuchen. Die Signifikanz der Ergebnisse in Tabelle 3 können von der (Wirtschafts-)Informatikforschung nicht ignoriert werden, da ein Anbieterausfall direkte Auswirkungen auf Datensicherheit und Servicebereitstellung hat. Daher müssten die bestehenden, theoretischen Rahmenwerke (siehe Tabelle 1) zwingend um die ***Dimension „Anbieterausfall“*** ergänzt werden. Diese Arbeit stützt sich auf das Berechnungsschema nach Altman, welches speziell für den amerikanischen Markt entwickelt wurde. Die Forschung der kommenden Jahre könnte bei einer entstehenden breiteren Datenbasis die Finanzberichte von Anbieteren aus zurückliegenden Jahren bewerten und mit echten Anbieterausfällen korrelieren, was sich derzeit aufgrund der stark begrenzten Anzahl an insolventen Anbietern, die börsennotierten sind, eher schwierig gestaltet. Ferner müssten neue Modelle konstruiert werden, die zum einen auf den CC-Markt zugeschnitten sind und zum anderen regionale Unterschiede berücksichtigen. Hierdurch entstünde der Praxis die erforderliche Unterstützung, die bei der Bewertung der CC-Anbieter notwendig ist. Für die Bewertung des deutschen Marktes könnte beispielsweise der Bundesanzeiger dienen, der Berichte unabhängig von der rechtlichen Form der Unternehmen veröffentlicht. Bei einer intensiven CC-Nutzung (z.B. für Kerngeschäftsprozesse) könnten spezielle Modelle zur ***Bewertung von Stresstests*** entwickelt werden, die den Nutzern ein geeignetes Risikomanagement bereitstellen. Darin sollten Szenarien und zeitliche Entwicklungen beachtet werden. Ferner sind weitere Forschungstätigkeiten im Bereich Vertrauenswürdigkeit von CC-Anbietern nötig, die neben den Ausfallrisiken ***auch andere Faktoren*** wie fehlende Zertifikate, fehlender Zugang zu Technologien, schwaches Management etc. beinhalten könnten. Auch die Einflüsse und

Auswirkungen der jeweiligen Faktoren untereinander stellen ein interessantes Forschungsfeld dar. Auf der Seite der CC-Anbieter gilt es um Vertrauen zu werben und die Nachhaltigkeit des Unternehmens unter Beweis zu stellen. Die Analyse von ***Signaleffekten in der IT*** stellt ein recht neues Forschungsfeld dar [BH11], welches auch um die Aspekte dieser Arbeit ergänzt werden kann. Nach Benlian und Hess [BH11] besitzen insbesondere diejenigen Indikatoren eine hohe Signalwirkung, die messbar und transparent sind. Beide Anforderungen können für die hier dargestellte Berechnung der Anbieterausfallrisiken als „erfüllt" bezeichnet werden. Für die IT-Praxis wird ein wichtiges Problem transparent, welches proaktiv gesteuert werden muss. Die Bewertung der Kreditwürdigkeit von diversen Zulieferern ist primär Aufgabe des Finanz- und Einkaufswesens. Durch die Nutzung von CC und der einhergehenden organisatorischen Veränderungen - „Making users to choosers" - müssten IT-Abteilungen künftig vermehrt gewisse ***Grundprinzipien aus dem Finanzwesen*** beherrschen. Zwar ist das IT-Outsourcing kein neues Phänomen. Aber die zunehmende CC-Nutzung sowie die Kleinteiligkeit und Intransparenz des CC-Markts erfordert ein stärkeres Bewusstsein für Ausfallrisiken. An dieser Stelle sei erwähnt, dass die Auditoren der Finanzberichte lediglich die Richtigkeit der Berichte prüfen und auf Risiken meist nur unzureichend hinweisen [Mc03]. Somit obliegt die qualitative Bewertung allein dem IT-Nutzer. Ferner gilt es CC-Anbieter nicht nur einmalig im Rahmen eines Selektionsprozesses zu bewerten, sondern periodisch auf ihre Ausfallrisiken zu kontrollieren. Bei offensichtlich kritischen CC-Anbietern müssten im Sinne eines ***aktiven Risikomanagements*** umso dringender Exit-Strategien und Alternativszenarien überlegt werden. Bereits die Verträge mit CC-Anbietern sollten „Disaster Recovery"-Szenarien angemessen adressieren, damit die Nutzer den CC-gestützten Geschäftsprozess zeitnah wieder aufnehmen können. Die Bandbreite dieser Sicherungsszenarien kann von einfachen back-up Lösungen bis hin zur Duplikation des gesamten CC-Services bei einem weiteren Provider reichen [FK15]. Ferner ist die Nutzung proprietärer CC-Software i.d.R. mit einer spezifischen Datenstruktur verbunden (Lock-in-Effekte) [Du13], was sich im Falle einer Anbieterinsolvenz umso problematischer gestaltet. Die Auswirkungen von Insolvenzrisiken beziehen sich nicht nur auf Endnutzer eines CC-Services sondern auch auf andere Teilnehmer im CC-Netzwerk. Wie eingangs erwähnt stehen hinter einem Service oftmals mehrere CC-Unternehmen, die stark voneinander abhängig sein können. Daher sollten Teilnehmer eines CC-Netzwerks ***neue Teilnehmer auf Ausfallrisiken*** testen und die Auswirkungen simulieren. Bezüglich der Ergebnisse in Tabelle 3 muss konstatiert werden, dass nur „reine" CC-Unternehmen berücksichtigt wurden, indem Anbieter mit diversifiziertem IT-Angebot bestehend aus CC und anderen Services im Rahmen der Z-Faktorenanalyse exkludiert wurden. Weitere Stichprobentests indizierten, dass vor allem die ***großen, namhaften Anbieter aufgrund der Diversifikation geringere Ausfallrisiken*** (z.B. SAP) aufweisen als kleinere, reine CC-Anbieter, was die Vertrauenswürdigkeit traditioneller Anbieter stärkt.

6 Limitationen und Fazit

Wie jede wissenschaftliche Arbeit, weist auch diese potentielle Limitationen auf. Dem-

nach besteht die Möglichkeit, dass nicht alle relevanten CC-Unternehmen in der Selektionsphase gefiltert wurden. Die Ursachen hierfür können in der Unvollständigkeit der gewählten Quellen oder in alternativen Bezeichnungen vermeintlicher CC-Anbieter liegen. Ferner wurden nur börsennotierte, überwiegend amerikanische CC-Unternehmen berücksichtigt, womit die Ergebnisse schwer auf deutsche Unternehmen mit anderen Rechnungslegungsstandards übertragbar sind. Jedoch muss festgehalten werden, dass alle selektierten CC-Anbieter auf dem deutschen Markt aktiv sind. Weiterhin ist die Grundgesamtheit mit 38 CC-Anbietern zwar größer als von Altman gefordert. Aber ein Rückschluss von der Empirie auf die Allgemeinheit ist zumindest nicht kritiklos möglich. Dennoch bietet diese Arbeit wichtige, neue Einblicke im Themengebiet „Vertrauen im CC-Markt“ und deckt die aktuellen Risiken von CC-Anbietern auf. Ferner geht diese Arbeit mit der Meinung von IT-Experten von Gartner einher [Ga15], die annehmen, dass jeder vierte CC-Anbieter das Jahr 2015 nicht überstehen wird. Dementsprechend groß sind auch die Auswirkungen auf die Daten- und Geschäftsprozesssicherheit der Nutzer. Der recht junge CC-Markt befindet sich wie jeder neue Markt in einer Findungsphase, bei der adäquate Selektionstechniken bei der Anbieterauswahl unabdingbar sind. Die künftige Forschung sollte sich daher insbesondere der Frage widmen, *wie* diese signifikanten Ausfallrisiken im CC-Selektionsprozess geeignet Beachtung finden können.

Literaturverzeichnis

[Al00] Altman, E.I.: Predicting financial distress of companies: revisiting the Z-score and ZETA models. Stern School of Business, New York University, 2000, S. 9-12.

[Al68] Altman, E.I.: Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The journal of finance, 1968, 23(4), S. 589-609.

[Ar10] Armbrust, M. et al.: A view of cloud computing. Communications of the ACM, 2010, 53(4), S. 50-58.

[At12] Atherton, A.: Cases of start-up financing: An analysis of new venture capitalisation structures and patterns. Int. J. of Entrepr. Behavior & Research, 2012, 18(1), S. 28-47.

[Be05] Bemmann, M.: Verbesserung der Vergleichbarkeit von Schätzgüteergebnissen von Insolvenzprognosestudien. No. 08/05. Dresden Discussion Paper in Economics, 2005.

[BH11] Benlian, A.; Hess, T.: The Signaling Role of IT Features in Influencing Trust and Participation in Online Communities. Int. J. of Elect. Commerce, 2011, 15 (4), S. 7-56.

[Du13] Dutta, A. et al.: Risks in enterprise cloud computing: the perspective of IT experts. Journal of Computer Information Systems, 2013, 53(4), S. 39-48.

[FK15] Favre, K.; Känzig, D.: IT and Internet Newsletter Switzerland: Legal Aspects of Cloud Computing, available at: http://thouvenin.com/wp-content/uploads/2015/01/20150126-IT-and-Internet-Thouvenin-Newsletter-Switzerland.pdf (Abruf am 18.03.2015).

[Fe06] Fettke, P.: State-of-the-Art des State-of-the-Art: Eine Untersuchung der Forschungsmethode “Review” innerhalb der Wirtschaftsinformatik. Wirtschaftsinformatik, 2006, 48(4), S. 257-266.

[Ga15] Gartner, One of four cloud provider will be gone by 2015, available at: http://www.computerworld.com/article/2486691/cloud-computing/one-in-four-cloud-providers-will-be-gone-by-2015.html (Abruf am 18.03.2015).

[Gr11] Grobauer, B., et al.: Understanding cloud computing vulnerabilities. Security & privacy, IEEE, 2011, 9(2), S. 50-57.

[KK14] Keller, R.; König, C.: A Reference Model to Support Risk Identification in Cloud Networks. Proc. of the 35th Int. Conf. on IS (ICIS), 2014, Auckland.

[Le10] Leimeister, S. et. al.: The Business Perspective of Cloud Computing: Actors, Roles and Value Networks. Proc. of the 18th Europ. Conf. on IS (ECIS), 2010, Pretoria.

[Ma11] Marston, S. et al.: Cloud computing - The business perspective. Decision Support Systems, 2011, 51(1), S. 176-189.

[MT12] Martens, B.; Teuteberg, F.: Decision-making in cloud computing environments: A cost and risk based approach. Information Systems Frontiers, 2012, 14(4), S. 871-893.

[Mc03] McKee, T.E.: Rough sets bankruptcy prediction models versus auditor signalling rates. Journal of Forecasting, 2003, 22(8), S. 569-586.

[Me12] Meer, D.V. et al.: A cost-based database request distribution technique for online E-commerce applications. MIS Quarterly, 2012, 36(2), S. 479-507.

[MG11] Mell, P.; Grance, T.: The NIST definition of cloud computing", available at: http://pre-developer.att.com/home/learn/enablingtechnologies/The_NIST_Definition_of_ Cloud_Computing.pdf (Abruf am 18. März 2015).

[MC13] Morgan, L.; Conboy, K.: Factors affecting the adoption of cloud computing: an exploratory study. Proc. of the 21st Europ. Conf. on IS (ECIS), 2013, Utrecht.

[Ol12] Olson, D.L. et al.: Comparative analysis of data mining methods for bankruptcy prediction. Decision Support Systems, 2012, 52(2), S. 464-473.

[Pr10] Provalis Research (2010). WordStat 6: Content Analysis Module for QDA Miner & SimStat. http://www.provalisresearch.com (Abruf am 13 Februar 2015).

[Ri14] Ridder, F.: Gartner Survey Shows One in Four German CIO Respondents Have Adopted Cloud and More Plan to Increase Cloud Investments, available at: http://www.gartner.com/newsroom/id/2731417 (Abruf am 07.02.2015);

[Ro98] Rousseau, D.M. et al.: Not So Different After All: A Cross-Discipline View of Trust. Academy of Management Review, 1998, 23(3), S. 393–404.

[Su01] Sullivan, D.: Document Warehousing and Text Mining", Wiley, New York, NY, 2001.

[Ta05] Tan, P.N. et al.: Introduction to Data Mining. Addison-Wesley, Boston, MA, 2005.

[Wa14] Walterbusch, M. et al.,: Hybride Wertschöpfung durch Cloud Computing. Dienstleistungsmodellierung. Springer Fachmedien Wiesbaden, 2014, S. 155-174.

[Wa13] Walterbusch, M. et al.: Exploring Trust In Cloud Computing: A Multi-Method Approach. Proc. of the 21th Europ. Conf. on IS (ECIS), 2013, Utrecht.

[WC10] Wang, Y.; Campbell, M.: Business failure prediction for publicly listed companies in China. Journal of Business and Management, 2010, 16(1), S. 75-88.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Ein Rahmenwerk für Datenschutz-Metriken in der Cloud

Sebastian Luhn[1], Maxmilian Hils[2]

Abstract: Eine automatisierte Überprüfung der Erfüllung von Datenschutzanforderungen in Cloudsystemen kann Kunden von Cloudanbietern helfen, notwendigen Kontrollpflichten einfacher nachzukommen und bietet zudem die Möglichkeit, diese Überprüfungen kontinuierlich zu vollführen. Zu diesem Zweck können Datenschutzmetriken verwendet werden, die anhand einer Vielzahl von Messungen die Eigenschaften des Cloudsystems bezüglich der Anforderungen bewerten. Diese Ausarbeitung stellt ein Rahmenwerk vor, mit dem solche Metriken konstruiert werden können. Es ist theoretisch in der Messtheorie fundiert und nutzt zudem das Konzept des Konstrukts aus den Sozialwissenschaften, um die oft komplexen und nicht direkt messbaren Anforderungen des Datenschutzes messbar zu machen. Außerdem berücksichtigt es domänenspezifische Anforderungen, die sich aus der Nutzung von Metriken in IT-Unternehmen ergeben. Am Beispiel einer Standortmetrik wird gezeigt, dass das Rahmenwerk zur Konstruktion komplexer Metriken genutzt werden kann, die Aussagen über die Erfüllung von Datenschutzanforderungen ermöglichen.

Keywords: Automatisierte Kontrolle, Datenschutz, Messquellen, Metrik, Verarbeitungsstandort.

1 Einleitung

Die Möglichkeit für Unternehmen, eigene IT-Infrastrukturen in die Cloud auszulagern, erfreut sich steigender Beliebtheit. Dennoch gibt es nach wie vor Bedenken und Hinderungsgründe, die gegen den Wechsel von internen zu externen IT-Infrastrukturen sprechen. Ein wesentlicher Aspekt sind Datenschutzbedenken und Anforderungen, die sich aus dem Datenschutzrecht ergeben. So müssen sich potenzielle Cloud-Kunden auf die Angaben des Cloud-Anbieters verlassen, was etwa den Speicherort der Daten angeht. Außerdem kann dieser häufig auch nicht während des Betriebs der Cloud überwacht werden, da viele Cloud-Anbieter keine (überprüfbaren) Angaben zu Speicherorten machen[3]. Eine Auditierung des Anbieters zur Überprüfung dieser Angaben ist zwar möglich und vorgeschrieben, findet aber üblicherweise nur punktuell statt [Bo12]. Eine kontinuierliche Überprüfung ist damit nicht möglich.

In diesem Artikel wird ein Rahmenwerk präsentiert, das die automatisierte Überprüfung der Einhaltung von Datenschutzanforderungen bei Cloud-Anbietern ermöglicht. Dies geschieht durch die Konstruktion von Datenschutzmetriken, die eine Vielzahl von Datenquellen aggregieren, um Aussagen über die Einhaltung von Datenschutzanforderungen

[1] Forschungsgruppe IT-Sicherheit, Universität Münster, Leonardo-Campus 3, 48149 Münster, sebastian.luhn@wi.uni-muenster.de
[2] Forschungsgruppe IT-Sicherheit, Universität Münster, Leonardo-Campus 3, 48149 Münster, maximilian.hils@uni-muenster.de
[3] Google bietet z. B. die Möglichkeit, den Speicherort bestimmter Daten zu spezifizieren. Ob die Daten tatsächlich dort liegen, kann nicht überprüft werden. Siehe `https://cloud.google.com/storage/docs/bucket-locations`

geben zu können. Metriken werden bereits, unabhängig von Cloud-Anwendungen, häufig in IT-Unternehmen eingesetzt [So11]. In der Literatur werden diese Art von Metriken aber sehr praxisnah behandelt (vgl. [So11, AS12, CA05]) Dieser Artikel befasst sich daher in Abschnitt 2 zunächst mit den theoretischen Grundlagen der Messtheorie und Konzepten zur Messung komplexer Phänomene, wie sie im Bereich der Datenschutzanforderungen üblich sind. Anschließend werden in Abschnitt 3 domänenspezifische Anforderungen und das übliche Vorgehen bei der Verwendung von Metriken in IT-Unternehmen vorgestellt. Es folgt die Beschreibung des Rahmenwerks und der Konstruktion von Metriken in Abschnitt 4. Zudem wird es für die Bestimmung des Standorts einer virtuellen Maschine beispielhaft in Abschnitt 4.2 angewendet, bevor Abschnitt 5 mit einer Zusammenfassung und einem Ausblick den Artikel beschließt.

2 Theoretische Grundlagen

Die theoretische Basis des hier vorgestellten Rahmenwerks stammt im Wesentlichen aus zwei wissenschaftlichen Disziplinen. Die Messtheorie beschäftigt sich mit der Messbarkeit verschiedener Aspekte der Realität und ist damit Grundlage aller Messungen. Vor allem in den Sozialwissenschaften ist jedoch häufig das Ziel, auch nicht direkt messbare Phänomene einer Größe zuzuordnen. Dies ist bei Datenschutzmetriken ebenfalls der Fall. Eine Datenschutzmetrik misst, inwieweit ein System, bspw. eine virtuelle Maschine in der Cloud, die Datenschutzanforderungen, die aus Gesetzen und Regularien stammen, erfüllt. Beispielsweise dürfen bestimmte, personenbezogene Daten nur innerhalb des Europäischen Wirtschaftsraums (EWR) gespeichert werden. Eine entsprechende Datenschutzmetrik dazu gäbe an, inwieweit diese Anforderung bezüglicher aller betroffenen Daten erfüllt wird. Die Konzepte des Konstrukts und des Indikators, die in Abschnitt 2.2 erläutert werden, bilden das Fundament für die Approximation dieser Phänomene.

2.1 Messtheorie

Grundlage jeden Messens ist die Messtheorie [SM11]. Sie beschäftigt sich mit der Frage, ob Aspekte bzw. Phänomene der Realität überhaupt gemessen werden können und ist somit wesentlicher Bestandteil der Wissenschaftstheorie. Während statistische Methoden dazu dienen, von Daten auf darin enthaltene Informationen zu schließen, beschäftigt sich die Messtheorie damit, die Realität in Form von Daten abbilden zu können. Eingeführt wurde die Messtheorie von Stevens [St46], wesentliche Beiträge stammen zudem von Suppes [Su14].

Es wird zwischen zwei verschiedenen Varianten der Messtheorie unterschieden, der repräsentativen und der operationalen Messtheorie. Die repräsentative Messtheorie wird dabei auch als die klassische Messtheorie bezeichnet [SM11]. Messen im Sinne dieser Theorie bedeutet, die *Struktur der Realität* durch Zahlen abzubilden, d. h., eine *Repräsentation* der (empirischen) Realität durch Zahlen zu erstellen. Insbesondere bedeutet die Abbildung der Struktur der Realität, dass Relationen realer Objekte auf die numerischen Re-

präsentationen abgebildet werden. Das empirische Objekt und die Messmethode sind dabei unabhängig voneinander und wichtige Aufgabe der Messtheorie ist es, zu zeigen, dass eine gewählte Methode dem empirischen Objekt gerecht wird. Die der repräsentativen Messtheorie entgegengesetzte *operationale* Messtheorie hingegen besagt, dass ein Objekt allein durch die verwendete Messmethode definiert wird und damit mit dieser eine Einheit bildet. Ein Bezug zur Realität ist nicht unbedingt vonnöten, da er nicht Teil der Betrachtung dieser Theorie ist [Ha96]. Ein operationales Vorgehen bedingt, dass sich Skalen entweder von selbst ergeben oder eine Auswahl mehrerer, eventuell äquivalenter Skalen nur schwer möglich ist [Ba00]. Die Logik hinter der operationalen Messtheorie ist, dass die Messmethode Teil des gemessenen Objekts ist – und damit auch die Skala. Bemisst man die Intelligenz eines Menschen etwa mittels eines IQ-Tests und basierend auf einer anderen Skala, so wären dies nach der operationalen Messtheorie zwei verschiedene Objekte.

Um Relationen zu überprüfen, werden Vergleiche vorgenommen. Die Anzahl möglicher Vergleiche ist jedoch je nach empirischem Objekt unterschiedlich. So kann bspw. die Größe zweier Objekte verglichen werden, die dann auch Aussagen über das Größenverhältnis zulassen. Bei anderen Objekten kann hingegen u. U. nur die Aussage getroffen werden, ob sie unterschiedlich oder gleich sind.

2.1.1 Skalenniveaus und Gütekriterien

Die Anzahl möglicher Vergleiche drückt das *Skalenniveau* einer Messung aus. Es gibt an, wie *eindeutig* ein Objekt mittels numerischer Werte beschrieben werden kann bzw. wie viel Information eine Skala enthält. Dabei wird zwischen Nominal-, Ordinal-, Intervall- und Verhältnisskalen unterschieden.

Die am wenigsten eindeutige Skala ist die *Nominalskala*. Auf dieser Skala gemessene Objekte werden in Kategorien unterteilt, wobei die einzige mögliche Aussage über zwei Objekte unterschiedlicher Kategorien ist, dass sich die Objekte unterscheiden. Ein Beispiel hierfür sind IP-Adressen, bei denen die Zahlen lediglich der Unterscheidung verschiedener Adressen dienen. Es folgt die *Ordinalskala*, die eine Ranganordnung der Objekte zulässt, d. h., die mathematischen Operatoren $>$ und $<$ sind als Vergleich zweier Objekte zulässig. Es kann jedoch keine Aussage darüber erfolgen, wie weit zwei Objekte auseinanderliegen. Das von einem Kunden eines Cloudanbieters gewünschte Datenschutzniveau kann etwa auf einer solchen Skala gemessen werden. Zwischen den Abstufungen „normal “, „„hoch“ „sehr hoch“ ist eine eindeutige Rangfolge erkennbar. Es kann jedoch keine Aussage darüber getroffen werden, wie groß der Unterschied zwischen den einzelnen Niveaus ist. Die Möglichkeit der Addition und Subtraktion von Werten erlaubt erst die *Intervallskala*. Zwischen Objekten, die auf dieser Skala gemessen werden, kann der absolute Abstand bestimmt werden. Nur bei Messungen auf *Verhältnisskalen* können Aussagen wie „doppelt so groß wie…“ getroffen werden. Der Unterschied zwischen beiden Skalen wird am Beispiel der Temperatur deutlich: Eine Temperatur in Grad Celsius kann nicht doppelt so hoch wie eine andere sein, da sie auf einer Intervallskala, d. h. ohne absoluten Nullpunkt, gemessen wird. Das Verhältnis zweier Längen kann hingegen angegeben werden.

Ein weiterer wichtiger Aspekt einer Messung und insbesondere darauf anwendbarer statistischer Methoden ist deren *Bedeutsamkeit* [SM11]. Sie ergibt sich aus den erlaubten Transformationen, der Umkehrung der Eindeutigkeit: Je eindeutiger eine Skala ist, d. h., je mehr Aussagen über das Verhältnis zweier Objekte möglich sind, desto weniger mathematische Transformationen sind erlaubt, ohne die Struktur der Realität zu verändern. Bei Ordinalskalen ist bspw. die Bezeichnung der einzelnen Ränge mit Zahlen wahllos, d. h., es ist unerheblich, ob die Ränge nun mit 0 bis 10 oder etwa mit -5 bis 5 bezeichnet werden. Bedeutsam ist eine Messung nur dann, wenn alle erlaubten Transformationen, bspw. die Verdoppelung jedes Werts, nichts an den Relationen der einzelnen Messobjekte ändern. Für anwendbare statistische Methoden, etwa die Berechnung des arithmetischen Mittels, gilt: Das Ergebnis der Methode muss identisch sein, wenn entweder zuerst die Messdaten mittels einer erlaubten Transformation verändert werden und darauf die Methode angewendet wird oder das gleiche umgekehrt geschieht. So ist z. B. das arithmetische Mittel für Ordinalskalen nicht erlaubt, weil dieses abhängig von der Wahl der Werte unterschiedliche Ergebnisse liefert.

Ein Hauptproblem bei der Anwendung der repräsentativen Messtheorie ist, die erlaubten Transformationen der gemessenen empirischen Objekte zu erschließen. Dies ist insbesondere dann der Fall, wenn komplexe Phänomene gemessen werden sollen, aber nicht klar ist, was *genau* gemessen wird. Ein Beispiel aus den Sozialwissenschaften hierfür ist die Bemessung der Armut. Zudem ist auch die strikte Auslegung des Begriffs Bedeutsamkeit unter Umständen problematisch: So kann etwa ein arithmetisches Mittel über ordinal skalierte Objekte empirisch durchaus eine sinnvolle Bedeutung haben, selbst, wenn dies mathematisch bzw. messtheoretisch nicht erlaubt ist. Zur Lösung des letzteren Problems wird der Begriff der Bedeutsamkeit oftmals nur als *numerisch bedeutsam* verstanden [Ni94]. Dadurch kann es allerdings den empirisch sinnvollen Methoden an theoretischer Fundierung mangeln [SM11].

Wichtig bei der Erstellung von Datenschutzmetriken ist jedoch auch die Frage, ob eine (falsche) Wahl der Skala das Messergebnis beeinflussen kann oder statistische Methoden als möglich suggeriert, die tatsächlich nicht angewendet werden sollten. In der Literatur wird hierfür ein pragmatischer Ansatz vorgeschlagen, der direkt von den empirischen Daten ausgeht [Ga75]: Mittels Messmethoden wird auf Skalen gemessen, die meist durch die Methoden selbst bedingt sind. Leistungen von Schülern, die mittels Schulnoten gemessen werden, bedingen z. B. eine Ordinalskala. Wenn nun die Anwendung verfügbarer statistischer Methoden auch nach erlaubten Transformationen, etwa der Verdopplung der gemessenen Werte, zum gleichen Ergebnis kommt, kann die Skala verwendet werden, da durch sie keine falschen Schlüsse gezogen werden können.

Für Messungen existieren eine Reihe von Gütekriterien, um zu bestimmen, ob sie sinnvoll sind und zu überprüfbaren Ergebnissen führen können [Hi06]. Messungen, die in Datenschutzmetriken verwendet werden, müssen vor allem die Kriterien der Reliablität und der Validität erfüllen. Messungen, die das erste Kriterium erfüllen, sind reproduzierbar, d. h. bei wiederholter Messung desselben Sachverhalts ergibt sich das gleiche Ergebnis. Das Kriterium der Validität besagt, dass die Messung auch das misst, was tatsächlich gemessen werden soll. Die Bemessung dieses Kriteriums kann u. U. schwierig sein, wenn, wie

im letzten Abschnitt erwähnt, gar nicht klar ist, was genau gemessen wird. Hierzu muss dann auf die im nächsten Abschnitt erwähnten Konzepte zurückgegriffen werden.

2.2 Konstrukte und Indikatoren

Wie bereits im letzten Abschnitt erwähnt, kann die Anwendung der repräsentativen Messtheorie insbesondere bei komplexen Phänomenen schwierig sein. Außerdem kann es von Interesse sein, Aussagen über Phänomene treffen zu wollen, bei denen eindeutig feststeht, dass sie nicht direkt messbar sind. Diese Phänomene finden sich vor allem in den Sozialwissenschaften, sie können aber auch, wie später gezeigt wird, für Datenschutzmetriken in der Cloud auftreten. Es bedarf daher einer Möglichkeit, auf messtheoretischer Basis Aussagen über komplexe, nicht direkt messbare Phänomene treffen zu können.

In den Sozialwissenschaften haben sich dafür die Konzepte des *Konstrukts* und des *Indikators* etabliert. Ein Konstrukt ist das Phänomen, über das eine Aussage getroffen werden soll, das aber nicht direkt messbar ist. Zu diesem Zweck werden Indikatoren gesucht, die in einem semantischen Zusammenhang mit dem Konstrukt stehen und messbar sind. Mehrere Indikatoren werden mittels statistischer Methoden kombiniert und geben so eine indirekte Aussage über das Konstrukt.

Dieses Konzept des indirekten Messens hat zur Folge, dass immer auch von der operationalen, nicht nur der repräsentativen Messtheorie ausgegangen werden muss [Bo10]. Das ist dadurch bedingt, dass das Konstrukt per Definition nicht direkt messbar ist und die Auswahl, Gewichtung und Kombination der Indikatoren – die Operationalisierung der Indikatoren – semantischen Überlegungen folgt. Operationalisierungen können auch statistisch überprüft werden. Dabei wird der Einfluss von Indikatoren auf das zu bestimmende Konstrukt mittels Testdaten überprüft. Bei diesen Testdaten ist das Ergebnis bekannt, d. h., man weiß beispielsweise, an welchem Standort sich Daten befinden, für die eine Reihe von Messungen ausgewählter Indikatoren erfolgen. Daraus kann der Einfluss der einzelnen Indikatoren auf das Ergebnis erschlossen werden.

3 Domänenspezifische Anforderungen

Nachdem bisher die theoretischen Ansätze aus der Messtheorie und den Sozialwissenschaften erläutert wurden, befasst sich dieser Abschnitt mit den domänenspezifischen Anforderungen, die sich aus dem Kontext der Datenschutzmetriken in der Cloud ergeben, für die das Rahmenwerk entwickelt wurde. Dazu wird zunächst Literatur aus dem IT-Bereich vorgestellt, der sich mit Metriken beschäftigt. Aus diesen ergibt sich in Kombination mit den theoretischen Ansätzen der Forschungsansatz, der mit dem Rahmenwerk verfolgt werden soll.

3.1 Verwandte Arbeiten

Sowa [So11] bietet einen umfassenden Überblick über die Definition, Entwicklung und Verwendung von Metriken im IT-Umfeld. Die Definition einer Metrik orientiert sich dabei stark am National Institute for Standards and Technology (NIST) [NI08] bzw. dem entsprechenden ISO-Standard 27004 [IS09]. Danach sind Metriken Werkzeuge zur Performanzmessung und Entscheidungsunterstützung, die diese Aufgabe durch Sammeln, Analyse von Daten und Ausgabe von Informationen, d. h. kontextbezogenen Daten, erfüllen. Diese eher allgemein gehaltene Definition kann als mit einer Zielgröße versehene Approximation des in Abschnitt 2.2 erwähnten Konzepts des Konstrukts angesehen werden. Sowa beschreibt daraufhin die regulatorischen Anforderungen, aus denen heraus sich Anwendungsszenarien für Metriken ergeben können. Es folgt die Erläuterung des Messplans, der für die automatisiert erfolgenden Messungen wichtig ist. Die Skalen, auf denen diese Messungen stattfinden, entsprechen denen aus Abschnitt 2.1. Allerdings bedingt der starke Anwendungsbezug, dass Skalen hier bereits durch die Messung oder die danach stattfindenden Verarbeitungsschritte schon vorgegeben sind. Implizit wird hier von einer operationalen Messtheorie ausgegangen, auch, wenn sie nie benannt wird. Eine weitere Anforderung, die von Sowa erwähnt wird, ist die Notwendigkeit, verschieden stark aggregierte Metriken für unterschiedliche Zielgruppen innerhalb eines Unternehmens zu erstellen. Diese können sich etwa im Hinblick auf Frequenz und Detailgrad unterscheiden. Abschließend wird von Sowa beispielhaft eine Liste möglicher Metriken angegeben. In Abschnitt 3.2 werden Unterschiede dieser Art von Metriken zu den bereits aufgeführten theoretischen Ansätzen erläutert.

Ammann und Sowa [AS12] definieren wichtige Begriffe, wie sie für Metriken in IT-Unternehmen verwendet werden. Alle Messungen werden auf dem *Untersuchungsgegenstand* ausgeführt. Dieser ist beispielsweise eine virtuelle Maschine oder ein Server in der Cloud. Messungen erfolgen nach einem *Messplan*, der angibt, wie oft welche Messung durchgeführt wird und wie das Ergebnis dieser Messung aussieht. Das Ergebnis der Messungen sind *Rohdaten*. Zur Verarbeitung dieser Rohdaten, bspw. zur Erstellung von Durchschnittswerten, werden *Instrumente* verwendet. Diese sind einfache mathematische Verfahren, die auf die Rohdaten angewendet werden. Teil von Metriken im IT-Umfeld sind *analytische Modelle*, die angeben, wie Metriken berechnet werden. Außerdem geben *Entscheidungskriterien* mit *Schwellwerten* an, ob die gemessenen Daten das Kriterium erfüllen oder nicht. Jede Metrik hat einen *Zielwert*, den die Metrik erfüllen soll und der sich aus externen Anforderungen ergibt. Besonders wichtige Metriken werden *Kennzahlen* genannt. Alle diese Begriffe finden in angepasster Form Anwendung im Rahmenwerk. Die Autoren stellen außerdem ein Bottom-Up- sowie ein Top-Down-Verfahren zur Entwicklung von Metriken vor. Ersteres geht dabei von verfügbaren Messwerten aus und entwickelt aus diesen ein Messmodell, das Zusammenhänge zwischen Prozessen innerhalb von Unternehmen darstellt, aus dem heraus sich die zu verwendenden Metriken ergeben. Für das Top-Down-Verfahren werden Ziele und dazugehörige Fragen bzw. mögliche Problemstellungen formuliert, aus denen sich Metriken zur Beantwortung herleiten lassen. Dieses Verfahren heißt *Goal-Question-Paradigma*. Beide Verfahren werden in abgewandelter Form auch für das Rahmenwerk verwendet.

Einen Überblick über qualitative Sicherheitsmetriken, die in IT-Unternehmen Verwendung finden, geben Chapin und Akridge [CA05]. Dabei werden auch mögliche Probleme bei der Aussagekraft einzelner Metriken erwähnt, etwa, ob viele oder wenige entdeckte Viren innerhalb eines Systems als gut klassifiziert werden können. Im Fokus stehen auch hier vor allem Standards und regulatorische Anforderungen, denen Metriken gerecht werden sollen.

3.2 Forschungsansatz

Die in Abschnitt 2 erwähnten Methoden befassen sich auf theoretischer Basis mit der Art und Weise, wie Messungen vorgenommen und nicht direkt messbare Phänomene approximiert werden können. Anwendungsbeispiele finden sich häufig in den Sozialwissenschaften. Diese Grundlagen sind auch für die Erstellung von Datenschutzmetriken, die ebenfalls auf Messungen basieren und auch nicht messbare Zielgrößen umfassen, notwendig. Ziel dieses Rahmenwerks soll es sein, den speziellen Anforderungen von Datenschutzmetriken in der Cloud gerecht zu werden. Die Art und Weise, wie Metriken im IT-Umfeld verwendet werden, wurde im letzten Abschnitt beschrieben. Dabei zeigt sich, dass Metriken in IT-Unternehmen häufig direkt messbare Werte sind, selbst, wenn die eigentlich erwünschte Information nicht direkt messbar ist. Es gibt kein explizites Konzept des Konstrukts (vgl. Abschnitt 2.2). Dass Metriken Aussagen über die Phänomene von Interesse treffen können, wird entweder implizit angenommen oder rein verbal erläutert. Eine Operationalisierung (vgl. Abschnitt 2.1), also die Auswahl und Gewichtung von Indikatoren, findet nicht statt.

Das im folgenden Kapitel vorgestellte Rahmenwerk definiert Metriken hingegen so, dass sie Konstrukte approximieren, d. h., dass die Operationalisierung Teil der Metriken selbst ist. Es bringt dadurch die theoretischen Ansätze aus der Messtheorie und den Sozialwissenschaften in die Domäne IT-bezogener Metriken. Diese sind notwendig, um Aussagen über die nicht direkt messbare Erfüllung von Datenschutzanforderungen zu ermöglichen, ohne diese Information manuell und nur implizit aus einer Vielzahl von Indikatoren zu erschließen.

4 Ein Rahmenwerk für Metriken

In diesem Abschnitt wird das Rahmenwerk für Datenschutzmetriken in der Cloud vorgestellt. Dafür werden zunächst die Begriffe und Verknüpfungen der Elemente des Rahmenwerks erläutert und dargestellt. Es folgt die Beschreibung der Umsetzung der theoretischen und domänenspezifischen Prozesse zur Konstruktion von Metriken und der Erzeugung von Indikatoren anhand eines Anwendungsbeispiels. Abschließend werden analytische Modelle vorgestellt, die bei der Approximation der Konstrukte durch Metriken verwendet werden.

4.1 Begriffe und Verknüpfungen

Vorrangiges Ziel des Rahmenwerks ist es, die in den vorherigen Kapiteln vorgestellten theoretischen Konzepte sowie die domänenspezifischen Anforderungen von Datenschutzmetriken in der Cloud zu verbinden. Der erste Schritt ist dabei die Definition der Begriffe der Metrikenkonstruktion sowie deren Verknüpfungen untereinander. Eine Übersicht ist in Abbildung 1 zu sehen.

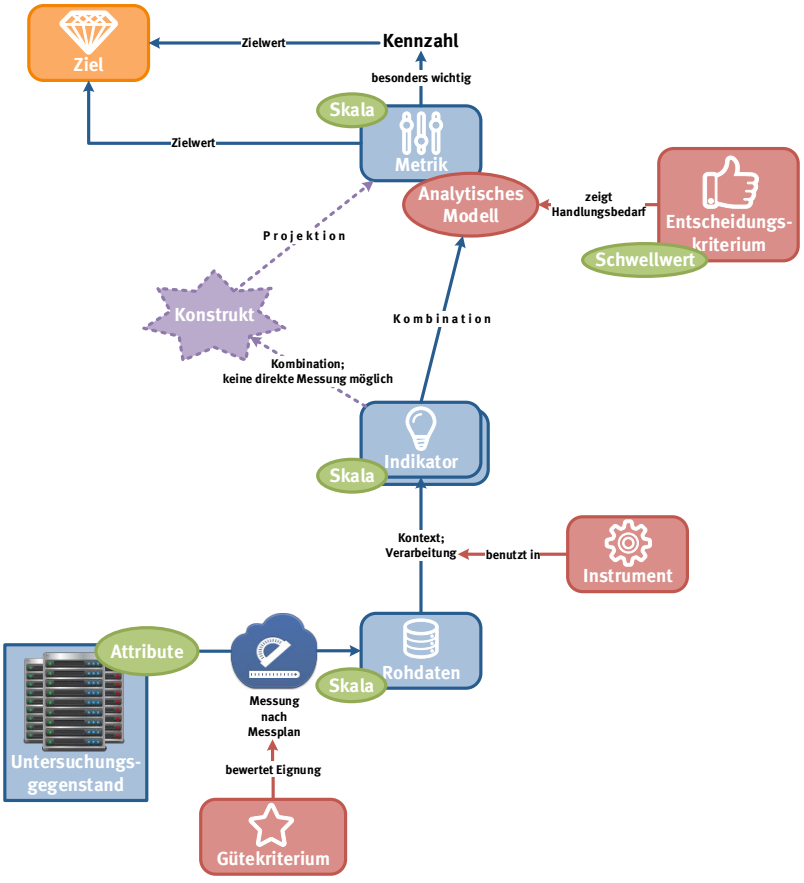


Abb. 1: Rahmenwerk für die Konstruktion von Metriken

Basis einer jeden Messung ist der *Untersuchungsgegenstand*. Dieser kann etwa eine virtuelle Maschine auf einem Server in der Cloud sein. Der Untersuchungsgegenstand besitzt Attribute, die mittels eines Messplans automatisiert gemessen werden können. *Gütekriterien* bewerten dabei die Eignung der Messung. Diese sind die bereits in Abschnitt 2.1 erwähnten messtheoretischen Kriterien wie Reliabilität, Reproduzierbarkeit, Validität und Generalisierbarkeit. Ergebnis der Messungen sind *Rohdaten*, die für die weitere Verarbeitung ge-

speichert werden. Diese Zusammenhänge ergeben sich sämtlich aus der in Abschnitt 3.1 erwähnten Literatur. Teile dieser Rohdaten können abhängig vom Kontext Indikatoren für ein zu approximierendes *Konstrukt* sein.

Das bereits in Abschnitt 2.2 eingeführte Konstrukt steht im Zentrum des Rahmenwerks und beschreibt die tatsächlich erwünschte Information. Gerade im Bereich der Datenschutzmetriken ist diese oft nicht direkt messbar, da sie sich üblicherweise aus datenschutzrechtlichen oder regulatorischen Anforderungen ergibt (vgl. Abschnitt 3.1). Ein Beispiel hierfür, die Bestimmung des Standorts einer virtuellen Maschine, wird in Abschnitt 4.2 behandelt. Aufgrund der nicht möglichen direkten Messungen muss das Konstrukt durch die Kombination messbarer Größen approximiert werden. Da die Konstrukte sich üblicherweise aus Anforderungen ergeben, geben die dazugehörigen Metriken den Erfüllungsgrad dieser Anforderung sowie einen Konfidenzwert an, der beschreibt, wie sicher die Aussage über den Erfüllungsgrad ist. *Ziel* einer Metrik ist, die Anforderung möglichst vollständig zu erfüllen. Die Notwendigkeit eines Zielwerts ergibt sich aus der vorgesehenen Anwendung des Rahmenwerks, der Messung der Erfüllung von Datenschutzanforderungen an Cloud-Dienste.

Metriken entstehen durch Kombination mehrerer *Indikatoren*. Die Art und Weise der Kombination bestimmt das *analytische Modell*, auf das in Abschnitt 4.3 näher eingegangen wird. Teil des analytischen Modells sind ein oder mehrere *Entscheidungskriterien*, mittels derer der Erfüllungsgrad der Metrik ermittelt wird.

Rohdaten, Indikatoren und Metriken werden jeweils auf – möglicherweise unterschiedlichen – Skalen gemessen (vgl. Abschnitt 2.1). Diese sind im Falle der Rohdaten und Indikatoren oft durch die Messungen selbst vorgegeben. Für die Metriken muss mittels des analytischen Modells (vgl. Abschnitt 4.3) eine passende Skala gefunden werden. Diese ergibt sich, wie in Abschnitt 2.2 erläutert, aus der operationalen Messtheorie bzw. pragmatischen Erwägungen.

Im nächsten Abschnitt wird das Vorgehen bei der Konstruktion von Metriken beispielhaft für die Bestimmung des Standortes einer virtuellen Maschine eines Cloud-Kunden dargestellt. Der Standort von Kundendaten bietet sich an, da rechtliche Anforderungen in diesem Bereich bestehen, etwa, dass personenbezogene Daten nicht außerhalb des Europäischen Wirtschaftsraums (EWR) gelagert werden dürfen. Zudem ist dieses Beispiel sehr anschaulich und es existieren eine Reihe von Indikatoren, sodass auch das analytische Modell einigermaßen umfangreich gestaltet werden kann.

4.2 Anwendungsbeispiel: Standortbestimmung

Metriken werden innerhalb des hier vorgestellten Rahmenwerks in einem kombinierten Top-Down-/Bottom-Up-Verfahren konstruiert. In diesem Abschnitt wird der Top-Down-Ansatz zur Konstruktion von Metriken aus Konstrukten heraus erläutert, der nächste Abschnitt befasst sich mit dem Bottom-Up-Verfahren zur Erzeugung von Indikatoren. Der kombinierte Ansatz ist notwendig, weil sich die Metriken zum einen aus den Anforderungen ergeben, die angeben, welche Eigenschaften des Untersuchungsgegenstands approxi-

miert werden sollen. Zum anderen ist eine Betrachtung der möglichen direkten Messungen notwendig.

Im Folgenden soll nun das Vorgehen der Metrikkonstruktion und die Anwendung des Rahmenwerks anhand eines Beispiels veranschaulicht werden. Als Ausgangspunkt liegt hier das Konstrukt des Speicher- und Verarbeitungsortes von Kundendaten, welches sich aus der datenschutzrechtlichen Anforderung, dass Daten nur in ausgewählten Ländern mit gesichertem Datenschutz gespeichert und bearbeitet werden dürfen, zugrunde. Es ergibt sich direkt aus den gesetzlichen Vorgaben („Wo werden die Daten gespeichert?“) und umfasst nicht etwa Mittel, mit denen diese Frage beantwortet werden könnte.

Auf Basis dieses Konstrukts lässt sich nun eine Metrik formulieren, die den Erfüllungsgrad der Anforderungen misst: „Alle Daten werden immer nur an erlaubten Standorten gespeichert.“. Zu dieser Metrik gilt es nun Indikatoren zu finden, die in einem semantischen Zusammenhang stehen.

Im Gegensatz zum bisherigen Top-Down-Verfahren vom Konstrukt zur Metrik werden die Indikatoren in einem Bottom-Up-Verfahren gebildet: Zentraler Untersuchungsgegenstand ist hier im Beispiel die virtuelle Maschine, von der aus mögliche Datenquellen ermittelt werden, die zu Indikatoren verarbeitet werden können. Beim Beispiel der Standortbestimmung lassen sich unter anderem die folgenden Daten nutzen:

- Bestimmung der minimalen Latenzzeiten zu nahegelegenen Servern zur Eingrenzung des physischen Standorts.
- Erfassung der bei Kommunikation mit Referenzservern verwendeten Kommunikationswege/Hops
- Generierung eines Fingerabdrucks der Maschinenumgebung (Andere Hosts im Netzwerk, MAC-Adressen, ...)
- Durchführung von DNS-Abfragen, welche je nach Standort unterschiedliche Antworten liefern

Diese Datenquellen gilt es in Anlehnung an das Rahmenwerk auf Informationsgehalt, Komplexität und Akzeptanz seitens des Cloudanbieters zu überprüfen. Beispielsweise lässt sich festhalten, dass die Generierung eines Fingerabdrucks zur Standortbestimmung nur begrenzt geeignet ist: Zum einen lassen sich mit einem Fingerabdruck nur Änderungen in der Netzwerkumgebung erkennen, diese können jedoch unabhängig von einer für die Metrik relevanten geographischen Änderung des Serverstandorts stattfinden. Zum anderen dürfte die Akzeptanz des Cloudanbieters für Scans der Maschinenumgebung vergleichsweise gering sein. Ziel der Indikatorenbildung ist jedoch gleichzeitig die Schaffung einer breiten Datenbasis, weshalb weniger aussagekräftige Indikatoren auch bei Vorhandensein von für die Metrik besseren/direkteren Indikatoren beibehalten werden sollen, auch um eine Kreuzvalidierung zu ermöglichen.

Im letzten Schritt sollen nun die verschiedenen Indikatoren miteinander kombiniert werden, sodass das daraus entstehende analytische Modell Aussagen darüber treffen kann, ob das Gesamtsystem die gestellten Anforderungen erfüllt. Wie anhand der beispielhaft

aufgeführten Datenquellen deutlich wird, ist eine Zusammenführung der Daten aufgrund unterschiedlicher Skalen nicht einfach.

4.3 Analytische Modelle

Wesentlicher Bestandteil einer Metrik ist das analytische Modell, mit dem die Indikatoren zu einem Wert kombiniert werden können. Die Auswahl des richtigen Modells, d. h. die in Abschnitt 2.2 erwähnte Operationalisierung der Indikatoren, beruht dabei hauptsächlich auf semantischen Überlegungen sowie der Validierung des gewählten Modells durch Testdaten.

Da sich die Konstrukte im Rahmen von Datenschutzmetriken für die Cloud aus datenschutzrechtlichen und regulatorischen Anforderungen ergeben, ist es Aufgabe des analytischen Modells, die gegebenen Ausprägungen der Indikatoren danach zu klassifizieren, ob das Gesamtsystem die Anforderungen erfüllt oder nicht. Wenn beispielsweise der Standort personenbezogener Daten in der Cloud bestimmt wird, gibt es, je nach den spezifischen Anforderungen, erlaubte und nicht erlaubte Standorte, an denen sich die Daten befinden können.

Allerdings gibt es keine einfache, eindeutige Möglichkeit der Bestimmung des Standortes. Daher werden eine Reihe direkter und indirekter Indikatoren für diesen Zweck herangezogen und in einem analytischen Modell kombiniert. Direkte Indikatoren geben dabei schon einen vermuteten Standort an, beispielsweise durch Datenbanken, die IP-Adressen Standorten zuweisen. Weitere Indikatoren werden benutzt, um die Zuverlässigkeit der Standortbestimmung zu erhöhen. Dabei werden auch indirekte Indikatoren benutzt. Diese geben selbst keinen direkten Hinweis auf den Standort, sondern werden als Fingerabdruck eines Standortes verwendet. Diese Indikatoren geben bspw. Auskunft über die an einem Standort in der Nähe befindlichen Geräte im gleichen Netzwerk, um diese mit vorher gewonnenen Fingerabdrücken zu vergleichen und einer Kategorie zuzuordnen.

Diese Zuordnung – sowie die spätere Bestimmung des Standorts auf Basis aller Indikatoren – geschieht mit statistischen Klassifikationsverfahren, etwa Entscheidungsbäumen und Nearest-Neighbours-Algorithmen. Gemein ist diesen Verfahren, dass, vereinfacht gesprochen, jede Sammlung von Messwerten (d. h., Messwerte aller Indikatoren zusammengefasst) mit bereits vorher in Kategorien unterteilten früheren Sammlungen verglichen wird und danach in die ähnlichste Kategorie sortiert wird. Entscheidungsbäume nehmen dafür Schwellwerte einzelner Indikatoren, die eine besonders prägnante Unterscheidung erlauben. Dieses Verfahren wird im Beispiel für die Klassifikation der indirekten Indikatoren verwendet. Bspw. könnte das Vorhandensein eines bestimmen Geräts im gleichen Netzwerk der virtuellen Maschine ein sehr prägnantes Unterscheidungsmerkmal sein. Beim Nearest-Neighbour-Verfahren wird diejenige Kategorie, d. h. derjenige Standort als Ergebnis geliefert, das die meisten ähnlichen Sammlungen von Messwerten aufweist wie die zu kategorisierende Sammlung. Das Ergebnis ist ein vermuteter Standort, der dann in einer Metrik daraufhin überprüft wird, ob er erlaubt oder nicht erlaubt ist.

5 Zusammenfassung und Ausblick

In diesem Artikel wurde ein Rahmenwerk für Datenschutzmetriken in der Cloud vorgestellt. Metriken finden bereits jetzt vielfältige Anwendung in IT-Unternehmen, hauptsächlich in den Bereichen IT-Sicherheit und Compliance, werden dort meist aber sehr praxisbezogen benutzt. Eine theoretische Fundierung des Messens ergibt sich aus der Messtheorie, die dabei hilft, die Messbarkeit empirischer Objekte zu untersuchen. Phänomene, die nicht direkt messbar sind, können mithilfe der Sozialwissenschaften, die das Konzept des Konstrukts für nicht direkt messbare Phänomene eingeführt haben, approximiert werden. Dieses Konzept ist für Datenschutzmetriken sehr nützlich, weil die Ziele dieser Metriken, die sich aus datenschutzrechtlichen und regulatorischen Anforderungen ergeben, oft nicht direkt messbar sind.

Der wesentliche Beitrag dieses Artikels ist die Kombination der theoretischen Ansätze mit der gängigen Praxis von Metriken im IT-Umfeld. Das Konstrukt als Repräsentation des nicht direkt messbaren Phänomens von Interesse ist hier zentraler Bestandteil. Mithilfe dieses Rahmenwerks ist es möglich, Metriken in einem kombinierten Top-Down-/Bottom-Up-Verfahren zu erstellen. Dies wurde am Beispiel der Bestimmung des Standorts einer virtuellen Maschine gezeigt.

Als nächster Schritt wird ein detailliertes analytisches Modell für die Standortbestimmung entwickelt, getestet und validiert werden. Damit kann dann gezeigt werden, inwieweit die hier präsentierte Herangehensweise an Metriken helfen kann, Datenschutzanforderungen in der Cloud automatisiert zu überprüfen.

Literaturverzeichnis

[AS12] Ammann, F.-E.; Sowa, A.: Systematische Entwicklung von Metriken zur Beurteilung der Datensicherheit. Datenschutz und Datensicherheit – DuD, 36(4):247–251, 2012.

[Ba00] Balzer, Wolfgang: Die Wissenschaft und ihre Methoden. Grundsätze der Wissenschaftstheorie. Journal for General Philosophy of Science / Zeitschrift für Allgemeine Wissenschaftstheorie, 31(1):179–186, 2000.

[Bo10] Bohrnstedt, George W: An Overview of Measurement in the Social Sciences. In: Presentation at the National Academy of Sciences Workshop on Advancing Social Science Theory: The Importance of Common Metrics, February. S. 24–25, 2010.

[Bo12] Borges, G. et al.: Datenschutzrechtliche Lösungen für Cloud Computing. Kompetenzzentrum Trusted Cloud, Oct 2012.

[CA05] Chapin, D. A.; Akridge, S.: How Can Security Be Measured. Information Systems Control Journal, 2:43–47, 2005.

[Ga75] Gardner, Paul Leslie: Scales and Statistics. Review of Educational Research, S. 43–57, 1975.

[Ha96] Hand, David J: Statistics and the Theory of Measurement. Journal of the Royal Statistical Society. Series A (Statistics in Society), S. 445–492, 1996.

[Hi06] Himme, Alexander: Gütekriterien der Messung: Reliabilität, Validität und Generalisierbarkeit. Methodik der empirischen Forschung, Wiesbaden, S. 383–400, 2006.

[IS09] ISO: ISO 27004:2009, Information Security Management — Measurement. International Organization of Standardization (Hrsg.), 2009.

[Ni94] Niederée, Reinhard: There Is More to Measurement than Just Measurement: Measurement Theory, Symmetry, and Substantive Theorizing. Review of Foundations of Measurement. Journal of Mathematical Psychology, 38(4):527–594, 1994.

[NI08] NIST: NIST 800-55: Performance Measurement Guide for Information Security. National Institute of Standards and Technology (Hrsg.), 2008.

[SM11] Saint-Mont, Uwe: Messtheorie. In: Statistik im Forschungsprozess, S. 23–76. Physica-Verlag HD, 2011.

[So11] Sowa, A.: Metriken, der Schlüssel zum erfolgreichen Security und Compliance Monitoring. 2011.

[St46] Stevens, S. S.: On the Theory of Scales of Measurement. Science, 103(2684), 1946.

[Su14] Suppes, Patrick: Foundations of Measurement, Jgg. 2. Elsevier, 2014.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Werkzeuge zur Messung der datenschutzkonformen Einhaltung des Verarbeitungsstandorts in der Cloud

Bernd Jäger[1] Reiner Kraft[2] Annika Selzer[3] Ulrich Waldmann[4]

Abstract: Automatisiert nutzbare Datenquellen zur Bestimmung des aktuellen Verarbeitungsstandorts können Benutzer eines Cloud-Dienstes darin unterstützen, ihrer Kontrollpflicht nachzukommen und das Vertrauen in eine datenschutzkonforme Verarbeitung ihrer Daten zu stärken. Dazu können Standortmetriken definiert werden, für die geeignete Datenquellen gefunden und sinnvoll kombiniert werden müssen, um sich dem komplexen Datenschutzziel wie dem zulässigen Verarbeitungsort anzunähern. Diese Ausarbeitung beschreibt die Vorgehensweise und prototypische Umsetzung einer solchen Standortmetrik. Jede ausgewählte Messquelle bietet allerdings nur Indizien, mit denen sich bestenfalls die Zuverlässigkeit einer Gesamtaussage erhöhen lässt. Mit Hilfe eines Prototyps sollen aussagekräftige und zuverlässige Datenquellen und Indikatoren für die Standortmetrik identifiziert werden.

Keywords: Automatisierte Kontrolle, Datenschutz, Messquellen, Metrik, Verarbeitungsstandort.

1 Herausforderung und Ziel

Für Unternehmen, die personenbezogene Informationen zu Kunden, Mitarbeitern oder anderen Personen in der Cloud verarbeiten lassen wollen, ist sehr wichtig, dass diese Daten dort hinreichend geschützt sind.[5] Ein Grund hierfür ist, dass die Unternehmen als Cloud-Nutzer für die in ihrem Auftrag in der Cloud verarbeiteten Daten im Rahmen der sogenannten Auftragsdatenverarbeitung verantwortlich bleiben. Um dieser Verantwortlichkeit gerecht werden zu können, legt der Gesetzgeber dem Auftraggeber einer Auftragsdatenverarbeitung[6] als verantwortliche Stelle die Pflicht auf, den Auftragnehmer[7] regelmäßig datenschutzrechtlich zu kontrollieren.[8] Die Kontrollpflicht im Rahmen der Auftragsdatenverarbeitung betrifft die durch den Auftragnehmer zum Schutz der von ihm verarbeiteten personenbezogenen Daten getroffenen technischen und organisatorischen Maßnahmen gem. § 9 BDSG, zum Beispiel Maßnahmen zum Zutritts- und Zugangsschutz für Räume und/oder Systeme, auf welchen personenbezogene Daten verarbeitet werden, Maßnahmen

[1] Colt Technology Services, Herriotstraße 4, 60528 Frankfurt, bernd.jaeger@colt.net
[2] Fraunhofer SIT, Schloss Birlinghoven, 53754 Sankt Augustin, reiner.kraft@sit.fraunhofer.de
[3] Fraunhofer SIT, Rheinstraße 75, 64295 Darmstadt, annika.selzer@sit.fraunhofer.de
[4] Fraunhofer SIT, Rheinstraße 75, 64295 Darmstadt, ulrich.waldmann@sit.fraunhofer.de
[5] Der Beitrag entstand im Projekt VeriMetrix, gefördert vom BMBF im Programm IKT 2020 – Forschung für Innovationen, Förderkennzeichen: 16KIS0053K. Teile des Beitrags wurden zudem von der Europäischen Union aus dem Europäischen Fonds für regionale Entwicklung und vom Land Hessen kofinanziert.
[6] Im Rahmen der Nutzung von Cloud-Diensten ist dies der Cloud-Nutzer.
[7] Im Rahmen der Nutzung von Cloud-Diensten ist dies der Cloud-Anbieter.
[8] Vgl. § 11 Abs. 2 Satz 2 und 4 Bundesdatenschutzgesetz – kurz: BDSG.

zur Sicherstellung der Verfügbarkeit personenbezogener Daten und Maßnahmen zum Eingabeschutz, die nachträglich überprüfbar machen, ob und von wem personenbezogene Daten in Datenverarbeitungssysteme eingegeben, verändert oder entfernt wurden. Eine solche Kontrolle ist jedoch im Cloud-Umfeld aufgrund des verteilten Rechnens und der oft großen geographischen Entfernungen problematisch. [Bo12], [Se13]

Ein Lösungsansatz besteht in der Anwendung von automatisch überprüfbaren Datenschutzmetriken, also Kennzahlen zur Beurteilung datenschutzrelevanter Eigenschaften auf Basis von vertrauenswürdigen Messdaten, mit welchen sich der Umsetzungsgrad von Datenschutzmaßnahmen des Cloud-Anbieters kontinuierlich anhand von interpretierbaren Indikatoren überprüfen lässt. Solche Überprüfungen könnten sowohl den Prüfumfang der technischen und organisatorischen Maßnahmen gem. § 9 BDSG als auch weitere datenschutzrechtlich relevanten Eigenschaften einschließen. In einer Umfrage bei über 200 kleinen und mittleren Unternehmen aus dem Gesundheitswesen, der Rechtsberatung und dem Finanzsektor[9] erwiesen sich neben der Kontrolle der technischen und organisatorischen Maßnahmen – hier insbesondere die Kontrolle des Zugriffsschutzes einschließlich der Verschlüsselung personenbezogener Daten sowie die Datentrennung bzw. Mandantentrennung – u. a. die Löschung bzw. die Sperrung[10] von Daten nach Beendigung des Vertrages mit dem Cloud-Anbieter sowie die Kontrolle der zulässigen Verarbeitungsstandorte als wichtige Inhalte für automatisierte Datenschutzkontrollen im Cloud-Umfeld.

Aus datenschutzrechtlicher Sicht ist – neben der Kontrolle der technischen und organisatorischen Maßnahmen – die Kontrolle des Verarbeitungsstandorts besonders bedeutsam, da hierfür zum Teil strenge Vorgaben existieren. Eine Verarbeitung personenbezogener Daten durch einen Auftragsdatenverarbeiter ist innerhalb des Europäischen Wirtschaftsraumes, kurz: EWR, [11] datenschutzrechtlich privilegiert und nach Abschluss eines Auftragsdatenverarbeitungsvertrages und der Durchführung regelmäßiger Kontrollen in der Regel zulässig. Die Zulässigkeit der Verarbeitung personenbezogener Daten außerhalb des EWR muss hingegen anhand einer zweistufigen Prüfung bewertet werden: Zunächst muss die Verarbeitung durch eine gesetzliche Erlaubnisnorm oder durch die Einwilligung aller Betroffenen legitimiert sein (erste Zulässigkeitsstufe), sodann muss – etwa durch umfangreiche Datenschutzverträge oder internationale Datenschutzabkommen – sichergestellt werden, dass die personenbezogenen Daten auch außerhalb des EWR ausreichend geschützt werden (zweite Zulässigkeitsstufe).[12] Diese Unterscheidung geht auf den Umstand zurück, dass personenbezogene Daten innerhalb des EWR einem hohen und weitestgehend einheitlichen Schutz unterliegen, der sich aus der Umsetzung der Europäischen

[9] Vgl. http://www-wordpress.sit.fraunhofer.de/verimetrix/wp-content/uploads/sites/11/2014/05/Verimetrix-Fragebogen.pdf

[10] Die Sperrung tritt an die Stelle einer Löschung, wenn einer Löschung z. B. gesetzliche Aufbewahrungsfristen entgegenstehen oder eine Löschung wegen der besonderen Art der Speicherung nicht oder nur mit unverhältnismäßig hohem Aufwand möglich ist. Vgl. §§ 20 Abs. 3, 35 Abs. 3 BDSG

[11] Der Europäische Wirtschaftsraum ist eine Freihandelszone zwischen der Europäischen Union, Island, Liechtenstein und Norwegen.

[12] Vgl. Düsseldorfer Kreis, Datenübermittlung in Drittstaaten erfordert Prüfung in zwei Stufenhttp://www.bfdi.bund.de/SharedDocs/Publikationen/Entschliessungssammlung/DuesseldorferKreis/12092013DatenuebermittlungInDrittstaaten.pdf

Datenschutzrichtlinie in den einzelnen Staaten ergibt.[13] Im Gegensatz dazu unterscheiden sich die Datenschutzsysteme sogenannter Drittstaaten – also Staaten außerhalb des EWR – sehr stark. [Se14]

Aufgrund der durchgängigen Virtualisierung der IT-Systeme in den unter Umständen weltweit verteilten Rechenzentren der Cloud-Anbieter bedeutet die zweifelsfreie Identifikation des Verarbeitungsstandorts eine hohe Herausforderung für die Entwicklung automatisierter Datenschutzkontrollen. In diesem Beitrag wird ein Lösungsvorschlag für diese Aufgabe beschrieben. Schwerpunkte sind dabei die Darstellung der Messverfahren zur Gewinnung valider Standortinformationen sowie die Skizze eines Architekturmodells für die sichere Erhebung, Verarbeitung und Speicherung der Messdaten. Abschließend werden einige Aspekte beschrieben, die bei der Weiterentwicklung der dargestellten Verfahren zu einem umfassenden Werkzeug zur kontinuierlichen Bewertung der Datenschutzqualität von Cloud-Angeboten zu berücksichtigen sind.

2 Verwandte Arbeiten

Die AG Rechtsrahmen des Technologieprogramms Trusted Cloud hat sich intensiv mit dem Problem der Kontrollpflicht beim Cloud-Computing auseinandergesetzt und schlägt in ihrem rechtspolitischen Thesenpapier [Bo12] eine vereinheitlichte Testatlösung zur Erfüllung der Kontrollpflicht vor. In dem Papier wird die Meinung vertreten, datenschutzrechtliche Kontrollen beim Cloud-Computing aufgrund des benötigten technischen und rechtlichen Spezialwissens von speziell zertifizierten Auditoren durchführen zu lassen, die für die Auditierung auch haften sollen. Zudem wird vorgeschlagen, die Kontrolle auf Basis einheitlicher, europaweit geltender Prüfkriterien durchzuführen. In [Se13] und [KNW13] wird dieser Vorschlag um eine automatisierte, kontinuierliche Datenschutzkontrolle auf Basis sicherer Log-Daten erweitert. Hierdurch wird es möglich, Datenschutzvorfälle zeitnah zu erkennen und auf diese zu reagieren.

Die Entwicklung von Metriken zur kontinuierlichen und weitgehend automatisierten Überprüfung der Einhaltung von Datenschutzanforderungen, ist eine Aufgabe, die sich überwiegend auf einem theoretischen und praktischen Neuland bewegt. Vorhandene Ansätze für Datenschutzmetriken – etwa das Mix-Modell von David Chaum oder andere Anonymitätsmetriken – sind auf das Problem der Anonymisierung personenbezogener Daten fokussiert, haben also keine umfassende Sicht auf den Datenschutz und wurden darüber hinaus bislang nur in geringem Umfang in der Praxis angewendet. Daneben gibt es vereinzelt Versuche, Datenschutzkennzahlen für organisatorische Sachverhalte zu erproben, etwa zur Anzahl und Bearbeitungsdauer von Datenschutzanfragen und -vorfällen.[14] Eine Reihe an Vorschlägen gibt es auch zu Informationssicherheitskennzahlen (siehe dazu z. B. die Publikationen [IS09, CI10, Ja07, NI08, So11]). Insofern Fragestellungen der IT- und Informationssicherheit betroffen sind, können diese auch in Datenschutzmetriken

[13] In Deutschland wurde die Europäische Datenschutzrichtlinie u. a. durch entsprechende Anpassungen im BDSG umgesetzt.

[14] Beispiele für solche Kennzahlen finden sich in der Telematikinfrastruktur der elektronischen Gesundheitskarte. Siehe `http://www.gematik.de/`.

einfließen. Allgemein anwendbare Metriken zur Überprüfung und Bewertung von Datenschutzeigenschaften existieren allerdings noch nicht.

Zur datenschutzfreundlichen Überprüfung des Verarbeitungsstandorts können die Ereignisse und Operationen der betreffenden virtuellen Maschinen (VM) analysiert werden. Die VM-Standorte sind dazu beispielsweise mittels Authentisierung der Rechenzentren bestimmbar [Ma11], was allerdings die aktive Mitwirkung der Cloud-Anbieter notwendig macht. VM-Standorte können mit anderen Verfahren auch betreiberunabhängig bestimmen werden. Beispielsweise können die Nutzer von VMs mittels aktiver Messungen von Paketumlaufzeiten zu einem Netzwerkknoten und wieder zurück („Round Trip Time", RTT) die virtuellen Koordinaten von umgebenen Netzwerkknoten und damit die relativen geographischen Positionen und Positionsänderungen von Cloud-VMs bestimmen. [Ri11] Zur Analyse von Standortdaten können verschiedenen Messdaten zu sogenannten Fingerprints zusammengefasst und mit statistischen Tools (z. B. mittels Programmiersprache R) und maschinellen Lernverfahren [Fl12] analysiert werden. Dabei werden aktuell gemessene Fingerprints mit vorliegenden Referenzdaten bekannter Standorte verglichen, ohne dass alle Teilinformationen der Fingerprints direkt verstanden oder interpretiert werden müssten. Solche heuristisch ermittelten Fingerprints wurden bereits zur Identifizierung von VM-Standorten eingesetzt. [Ri09], vgl. auch [JSW15]

3 Metriken und Messpunkte

Zur Entwicklung aussagekräftiger und realistischer Datenschutzmetriken sind diese zunächst aus den datenschutzrechtlichen Anforderungen abzuleiten. Im Anschluss an diese Top-Down-Betrachtung ist bottom-up zu prüfen, welche Datenquellen zur Verfügung stehen, um diese Metriken zu befüllen. Darauf aufbauend sind abschließend die Berechnungsverfahren der Metriken zu beschreiben. [JSW15]

3.1 Top-Down-Ableitung der Metriken

Im ersten Schritt zur Entwicklung von Metriken geht es darum, die bestehenden rechtlichen Anforderungen zu identifizieren und zu diesen passende Metriken zu finden. [LH15] Rechtliche Anforderungen ergeben sich u. a. aus dem Bundesdatenschutzgesetz, wo beispielsweise geregelt ist, dass eine privilegierte Auftragsdatenverarbeitung, bei der keine gesetzliche Erlaubnisnorm oder die Einwilligung der Betroffenen benötigt wird, geographisch auf den EWR beschränkt ist.[15] Eine Übermittlung personenbezogener Daten an Cloud-Anbieter außerhalb des EWR unterliegt demgegenüber wie oben beschrieben weiteren Anforderungen.

Um die beschriebene rechtliche Anforderung zu konkretisieren, benötigt es einer Darstellung der Maßnahmen, mit denen sie erfüllt werden können. Im Falle der geographischen Beschränkung des Auftragsdatenverarbeitungsprivilegs wäre eine solche Maßnahme z. B.,

[15] Vgl. § 11 BDSG i. V. m. § 3 Abs. 8 Satz 3, Abs. 4 BDSG.

die Datenverarbeitung personenbezogener Daten außerhalb des EWR zu unterbinden. Aus diesen Maßnahmen ergibt sich wiederum das sogenannte Konstrukt, das eine abstrakte, nicht direkt messbare Beschreibung des Messziels – z. B. die Kenntnis aller Verarbeitungsstandorte – darstellt. [LH15] Unter Berücksichtigung dieses Ansatzes lässt sich z. B. die folgende Metrik definieren:

$$\text{Prozentzahl zulässiger Standorte (PzS)} = \frac{\text{Anzahl zulässiger Standorte}}{\text{Gesamtanzahl der Standorte}} \times 100$$

In die oben angeführte Standortmetrik fließen in doppelter Weise kundenspezifische Anforderungen ein: Zum einen bestimmen diese, wann ein Standort als zulässig gilt, zum anderen legen sie auch fest, wie genau die Standortbestimmung erfolgen muss, um Verletzungen der Datenschutzanforderungen eines Kunden mit relativ hoher Wahrscheinlichkeit erkennen zu können.

Übliche Kundenanforderungen können beispielsweise sein, dass sich Daten nur innerhalb Deutschlands, des EWR oder eines Landes, das nach Einschätzung der Europäischen Kommission ein vergleichbares Datenschutzniveau bietet, befinden dürfen. Wieder andere Kunden erlauben darüber hinaus auch eine Datenverarbeitung in so genannten Drittstaaten[16], solange ein vergleichbares Datenschutzniveau durch entsprechende Verträge oder internationale Abkommen[17] sichergestellt ist. Im Extremfall ist eine Verarbeitung bei einem externen Cloud-Anbieter – wenn überhaupt – nur in einem räumlich streng festgelegten und sowohl physisch als auch netztechnisch vom übrigen Teil des Rechenzentrums abgetrennten Bereich zulässig.

Diese Beispiele machen deutlich, dass die Verfahren zur Standortermittlung mehrere Stufen und Kategorien bedürfen, um den verschiedenen Arten an Standortanforderungen gerecht zu werden. Neben der regionalen Verortung ist auch der Betreiber des Rechenzentrums zu bestimmen, das als Ort der Datenverarbeitung ermittelt wurde. Da bei hohen Anforderungen jeglicher Standortwechsel ausgeschlossen sein kann, muss ferner ein solcher Wechsel mit hoher Wahrscheinlichkeit detektiert werden können.

3.2 Bottom-Up-Bestimmung der Datenquellen

In diesem Schritt geht es darum, in der Infrastruktur eines Cloud-Dienstes aussagekräftige Datenquellen und daraus abgeleitete Indikatoren und Messverfahren zu finden, anhand derer die Top-Down abgeleiteten Metriken und damit der Umsetzungsgrad der Datenschutzanforderungen berechnet werden können. Bei der Auswahl dieser Datenquellen und Messverfahren ist darauf zu achten, dass sie sich nicht nur funktional eignen und vertrauenswürdige Daten liefern, sondern auch tatsächlich einsetzbar sind – hierfür kann es technische, organisatorische und rechtliche Barrieren geben. Beispielsweise kann die Nutzung einer Datenquelle sich verbieten, weil sie datenschutzrechtlich bedenklich ist, oder aber

[16] Als Drittstaaten werden im Datenschutzrecht diejenigen Staaten bezeichnet, die nicht dem EWR angehören.
[17] Hierzu zählen u. a. die EU-Standardvertragsklauseln, Binding-Corporate-Rules und das Safe Harbor Abkommen.

an der mangelnden Akzeptanz eines Cloud-Anbieters scheitern, der Störungen in seiner Cloud-Infrastruktur befürchtet.

Das Bottom-Up-Vorgehen zur Identifikation von Indikatoren und Messverfahren und deren Zuordnung zu den top-down formulierten Metriken soll nachfolgend am Beispiel der oben genannten Metrik zum Standort der Datenverarbeitung veranschaulicht werden. Diese Metrik erfordert konsequenterweise eine kontinuierliche Messung des Verarbeitungsstandorts der Daten eines Cloud-Nutzers, um die Wahrscheinlichkeit zu erhöhen, selbst zeitweilige Verstöße gegen dessen Vorgaben zu entdecken. Dies bedeutet im Cloud-Umfeld, dass insbesondere der physische Standort des Hosts identifiziert werden muss, auf dem die virtuellen Maschinen (VM), welche die Nutzerdaten verarbeiten, gestartet wurden. Dieser „physische Standort" der VMs ist besonders interessant, weil mit dem heutigen Stand der Technik zwar Daten auf dem Transportweg oder im dem Moment, da sie auf der virtuellen Festplatte gespeichert werden, verschlüsselt werden können, zur Verarbeitung durch den virtuellen Prozessor der VM jedoch im Klartext vorliegen müssen. Am „physischen Standort" der VM besteht also ein erhöhtes Risiko, dass auf sensible Daten unberechtigt zugegriffen wird.

Da die für Cloud-Anwendungen charakteristische Virtualisierung eine eindeutige Identifizierung des physischen Standorts signifikant erschwert, ist eine Mehrzahl an Indikatoren und Messverfahren erforderlich, um diesen mit einer zumindest sehr hohen Wahrscheinlichkeit bestimmen zu können. Diese Vielzahl erhöht darüber hinaus auch die Manipulationssicherheit und Zuverlässigkeit der Messungen. Zum einen wird es so für einen Cloud-Anbieter aufwändiger und damit unattraktiv, die Datenbasis in ihrer Gesamtheit zu manipulieren, zum anderen kann so der Ausfall eines Messverfahrens leichter kompensiert werden.

Eine wesentliche Leitidee bei einem Großteil dieser Messverfahren ist es, zu prüfen, ob die ermittelten Informationen zur Umgebung der virtuellen Maschinen mit bekannten und in einer Datenbank eingetragenen Mustern übereinstimmen, die für einen bestimmten Standort charakteristisch sind. Diese Muster bilden einen sogenannten „Fingerprint" eines Standorts.[18] Um den „Fingerprint" mit der VM zu verknüpfen, werden die Messprogramme („Agenten") innerhalb der gleichen VM gestartet, auf der auch reguläre Anwendungen Nutzerdaten verarbeiten werden (Datenverarbeitungs-VM). Um Messungen, die nicht aus einer VM selbst heraus durchgeführt werden, mit dieser zu verknüpfen, kann eine externe Messinstanz z. B. eine eindeutige Agenten-ID abfragen (etwa durch einen Netzwerk-Trace von „außen" mit Übertragung der Agenten-ID). Eine Verknüpfung kann aber auch indirekt über geeignete Umgebungsparameter, die der VM und der externen Mess-Instanz bekannt sind, hergestellt werden.

Geeignete Umgebungsparameter zur Zusammenstellung von Fingerprints sind in verschiedenen Bereichen typischer Cloud-Plattformen zu finden und können innerhalb und außerhalb der Infrastruktur des Cloud-Anbieters angesiedelt sein. Es sind sowohl aktive Messungen mit Diagnosewerkzeugen (Security Scannern etc.) als auch passive Messungen

[18] Bei diesen handelt es sich allerdings nicht um kryptographische Hashwerte, selbst wenn sie eine ähnliche Funktion haben wie die Fingerprints, die für die Verifikation kryptographischer Schlüssel verwendet werden.

(z. B. von vorhandenen Datenströmen mit Hilfe eines Sniffers) möglich. Die Bereiche der virtuellen Maschinen und des VM-Managements mit der darunter liegenden Hardware bieten in ihren Implementierungsdetails hilfreiche Anhaltspunkte für eine Bestimmung des Verarbeitungsortes, insoweit der Cloud-Anbieter den Zugriff auf solche Daten zulässt.

Mögliche Messungen im Bereich der Infrastruktur eines Rechenzentrums betreffen auch die nicht-öffentlichen Schnittstellen und zusätzliche von außen nicht sichtbare Informationen, die aus gezielten Abfragen von u. a. Management-Schnittstellen, internen Switches und Netzwerkkomponenten gewonnen werden können. Mögliche Messdaten, die als Parameter in die Berechnung eines Fingerprints einfließen können, sind u. a. auch IP-Adressen, Paketlaufzeiten und Paketrouten sowie Software-Versionen von Gerätetreibern, Virtualisierungskomponenten oder deren Management-Schnittstellen (insoweit verfügbar). Beispielsweise kann in einigen Fällen aus einer gemessenen IP-Adresse per „Reverse DNS Lookup“ der Hostname einer Cloud-Anbieter-Netzwerkdienstinstanz ermittelt und als Indikator für die globale Region zur Metrikenentwicklung herangezogen werden. [JSW15]

Die Messdaten werden von einer externen Komponente ausgewertet und zu charakteristischen Fingerprints zusammengestellt mit dem Ziel, sie bestimmten Regionen und Standorten zuzuordnen und als Indikatoren für die Standortmetrik zu interpretieren. Ist der Standort eines Fingerprint mit hoher Wahrscheinlichkeit bestimmt, kann er in Form von Referenzdaten zur Standortbestimmung anderer Fingerprints zugelassen werden.

3.3 Beschreibung der Berechnungsverfahren

Das Zusammenwirken der verschiedenen Einzelindikatoren in einer Metrik wird mit Hilfe eines sogenannten analytischen Modells beschrieben. Dieses Modell liefert die Berechnungsvorschrift für die Metrik und drückt aus, welche Relevanz jedes einzelne Messverfahren für die Approximation eines betrachteten Konstrukts hat. Damit legt es auch fest, wie aussagekräftig die Metrik insgesamt ist. Nachfolgend wird ein solches Modell für die angeführte Standort-Metrik beschrieben, vgl. Beitrag [LH15] im Rahmen dieses Workshops.

Diese Metrik ergibt sich (logischerweise) als Funktion der Einzelbewertungen zu den insgesamt betrachteten Standorten. Für die Prüfung der Zulässigkeit eines einzelnen Standorts können die folgenden, mit Messinstrumenten innerhalb und außerhalb der virtuellen Infrastruktur des Cloud-Kunden (bzw. des Cloud-Anbieters) gewonnenen Indikatoren dienen:[19]

- Ind-1: mithilfe von DNS-Anfragen und der Abfrage von Standortdatenbanken ermittelte Standorte öffentlicher Server,
- Ind-2: mithilfe von Traceroute-Abfragen, die innerhalb der VM abgesetzt wurden, ermittelte Knotenpunkte und Zielstandorte von Datenpaketen,

[19] Die genannten Indikatoren und Messmethoden sollen das Verfahren veranschaulichen. Für die Standortbestimmung können zahlreiche weitere Indikatoren und Messverfahren sinnvoll sein.

- Ind-3: mithilfe von ping-Anfragen, die außerhalb der VM abgesetzt wurden, ermittelte Zielstandorte von Datenpaketen,
- Ind-4: mithilfe eines Netzwerk-Sniffers zusammengestellte Informationen über die Geräte in der Umgebung einer virtuellen Maschine,
- Ind-5: mithilfe spezialisierter Detection-Tools[20] zusammengestellte Informationen über die verwendete Virtualisierungssoftware.

Für die Bestimmung des Standorts einer virtuellen Maschine werden diese Indikatoren in zwei unterschiedliche Analyseverfahren gruppiert: Zum einen in solche, die sich auf einer direkten Messung des Standorts stützen (Ind-1, Ind-2 und Ind-3), zum anderen in solche, bei denen ein Standort indirekt mithilfe einer Fingerprint-Datenbank und statistischen Klassifikationsverfahren[21] bestimmt wird (Ind-4 und Ind-5). Alle Verfahren unterstützen zunächst einmal die Standortidentifikation. Für die Zulässigkeit des gefundenen Standorts ist dieser, wie oben beschrieben, zusätzlich noch an den kundenspezifischen Anforderungen zu spiegeln.

Je größer die Übereinstimmung der Ergebnisse aus Direktmessungen und Fingerprint-Vergleichen ist, desto eindeutiger ist auch die Standortbestimmung. Im Idealfall – positive Ergebnisse bei allen Einzelindikatoren – kann der Standort mit einer hohen Wahrscheinlichkeit korrekt bestimmt werden und hat auch die daraus abgeleitete Bewertung seiner Zulässigkeit einen hohen Grad an Plausibilität. Umgekehrt gilt dies auch, wenn alle Einzelindikatoren darauf hinweisen, dass der Standort einer Virtuellen Maschine nur unzureichend bestimmt werden kann.

Schwieriger fällt eine solche Festlegung immer dann, wenn die Einzelindikatoren unterschiedliche Ergebnisse liefern. Sind die Ergebnisse aus der Direktmessung negativ, kann mit sehr hoher Wahrscheinlichkeit ein korrekt bestimmter Standort vermutet werden. Ist hingegen der Fingerprint-Vergleich negativ, kann bei gleichzeitigem positiven Befund aus den Direktmessungen der Konfidenzwert[22] des verwendeten Klassifikationsverfahrens eine Entscheidung unterstützen.

Idealerweise sind die Berechnungsverfahren zur Standortbestimmung so offen gestaltet, dass Änderungen in den für einen Standort charakteristischen Angaben unmittelbar nachvollzogen werden können. Darüber hinaus sollten sie flexibel an neue Verfahren (z. B. einen besseren Algorithmus zum Fingerprint-Vergleich) angepasst werden können.

4 Architektur

Anwender des VeriMetrix-Systems sind insbesondere Cloud-Kunden, die kontinuierlich prüfen möchten, ob und inwieweit ihre Datenschutzanforderungen aktuell erfüllt werden.

[20] z. B. Scoopy

[21] Statistikwerkzeuge wie R oder SPSS bieten hierfür ein breites Spektrum an Methoden an.

[22] Dieser Wert, ein Maß für die Vertrauenswürdigkeit einer Aussage, wird üblicherweise von den statistischen Werkzeugen geliefert

Eine Architektur zur Messung des Verarbeitungsstandorts sollte dies berücksichtigen und ferner, dass der Cloud-Anbieter und der Cloud-Kunde in Bezug auf die Messungen unterschiedliche Interessen haben können. Der Betreiber hat nur geringes Interesse, unbearbeitete Messdaten aus seiner Cloud-Infrastruktur an externe auswertende Komponenten zu übertragen, und fürchtet deren unkontrollierte Verbreitung. Zudem könnte er daran interessiert sein, kritische Daten auszufiltern oder gar zu manipulieren. Der Cloud-Kunde möchten unverfälschte Informationen über die beauftragte Datenverarbeitung erhalten, aber nutzerspezifische Daten möglichst nicht mit anderen Kunden des Cloud-Betreibers teilen.

Abbildung 1 zeigt die wichtigsten Komponenten des VeriMetrix-Systems, links die Komponenten auf Seiten des Cloud-Betreibers. Dieser betreibt kundenspezifische VMs, in denen die eigentliche Datenverarbeitung stattfindet. Messdaten sogenannter Referenz-VMs (links unten), deren Standorte bekannt sind, dienen dem VeriMetrix-System zum Vergleich mit aktuellen Messdaten der Kunden-VMs. In der Mitte und rechts zeigt die Abbildung einige Komponenten des Kunden und ggf. auch eines unabhängigen VeriMetrix-Betreibers, der mit der Erfassung und Auswertung von Messdaten beauftragt werden könnte („Trust-as-a-Service“). Auf Seiten des Cloud-Betreibers sind in den Kunden-VMs und Referenz-VMs betreiberunabhängige Messmodule integriert, welche die in Kapitel 3.2 beschriebenen Umgebungsparameter messen. Die zentrale Komponente zur Berechnung der Metriken ist der Kollektor in der Mitte des Bildes. Er unterhält weitere Messmodule, die außerhalb der Infrastruktur des Cloud-Anbieters Standortparameter erfassen können und greift zur Berechnung der Metriken auf eine Referenzdatenbank zu, die regelmäßig aktualisiert und erweitert wird. Der Kollektor bietet eine Schnittstelle für Auditoren, die z. B. Ergebnisse aus vor-Ort-Prüfungen manuell eintragen können, und Schnittstelle für den Kunden und seine Anwendungen.

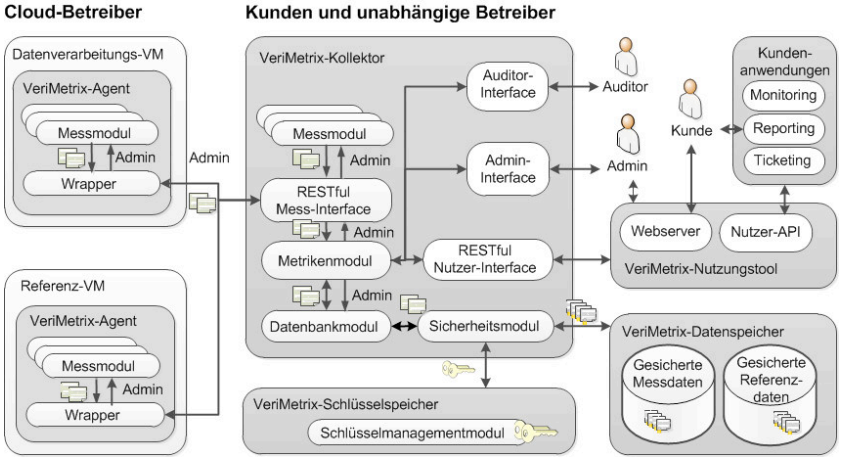


Abb. 1: VeriMetrix-Architektur

Ein Ziel dieses VeriMetrix-Systems ist, mutmaßliche oder tatsächliche Verstöße gegen Datenschutzanforderungen, etwa unzulässige Änderungen des Verarbeitungsstandorts, zuverlässig zu erkennen und ausreichend belegen zu können. Die Möglichkeit zur Konfiguration entsprechender Meldungen und zur Einsichtname in die einer Meldung zugrunde liegende Datenbasis sind daher wichtige Features der Nutzerschnittstelle.[23] Eine Administrationsschnittstelle auf Seiten des Auditors konfiguriert und steuert die VeriMetrix-Messmodule, beispielsweise den Umfang und die Häufigkeit von Messungen.

Die Frage, ob und wie die Messdaten geschützt werden müssen, betrifft nicht nur den Datenschutz, sondern auch die allgemeine IT-Sicherheit, da bekannt gewordene VM-Standortdaten auch als Grundlage für Angriffe dienen können.[24] [BS14] Wenn beispielsweise Umgebungsinformationen bestimmter VMs in die Hand von Angreifern fielen, könnten diese Angriffs-VMs mit höherer Erfolgsaussicht in derselben Umgebung starten. [Ri09] Die Architektur sieht daher Sicherheitskomponenten vor, welche die VM-spezifischen Messdaten in ein gesichertes Format überführen. In der potenziell unsicheren Umgebung des Cloud-Anbieters setzt das in Abbildung 1 gezeigte Sicherheitsmodul ein kryptographisches Verfahren ein, mit dem die Authentizität, Integrität und Vertraulichkeit der Messdaten geschützt werden. [KNW13]

Die Messdaten werden auf Basis eines von Schneier und Kelsey definierten Verfahrens gesichert, das Forward Integrity selbst in Umgebungen ermöglicht, denen nicht vorbehaltlos vertraut wird [SK99]. Unter dieser Voraussetzung können die Daten des VeriMetrix-Datenspeichers wiederum beim Cloud-Betreiber gespeichert werden, ohne dass der Cloud-Betreiber diese lesen oder verändern kann. Das Verfahren sieht vor, dass jeder neue Messwert mit dem vorigen Eintrag über Verschlüsselung, Hash-Werte und Prüfsummen verkettet und in eine aktuelle Messdatei geschrieben wird. Jedes Integritätsmodul handelt mit dem VeriMetrix-Schlüsselspeicher für jede neue Messdatendatei zwei symmetrische Startschlüssel aus. Das Sicherheitsmodul leitet daraus bzw. aus den Nachfolgeschlüsseln in jeder Runde für den aktuell zu erstellenden Messdateneintrag zwei symmetrische Arbeitsschlüssel für den zu erstellenden Eintrag ab: Durch Hashen der vorigen Arbeitsschlüssel zusammen mit den Zugriffsrechten für den neuen Eintrags werden daraus ein individueller Schlüssel zur Verschlüsselung der originären Messdaten und ein Schlüssel zur Berechnung einer Prüfsumme erzeugt. Nach jeder Verwendung und dem Erzeugen neuer Arbeitsschlüssel werden im Sicherheitsmodul die alten Schlüssel gelöscht. Die Startschlüssel stellen die Sicherheitsanker jeder gesicherten Datei dar und verlassen niemals den Schlüsselspeicher. Die Arbeitsschlüssel gesicherter Einträge sind nirgendwo gespeichert und können ausschließlich im Schlüsselspeicher aus den Startschlüsseln und den kodierten Zugriffsrechten rekonstruiert werden. Der Schlüsselspeicher sendet auf Anfrage diese Arbeitsschlüssel an autorisierte externe Instanzen (z. B. an einen Auditor), damit diese die Messdaten verifizieren, entschlüsseln und auswerten können.

[23] Aus wettbewerbsrechtlicher Sicht ist bei der zur Verfügung Stellung der Kontrollergebnisse darauf zu achten, Schmähkritik und nicht beweiskräftig nachweisbare Tatsachenbehauptungen zu unterlassen, keine offensichtlich unrichtigen Ergebnisse zu veröffentlichen sowie offenzulegen, wenn eine Bewertung ausschließlich auf den Angaben eines Cloud-Anbieters beruht. Vgl. analog das Ritter Sport-Urteil des Oberlandesgerichts München vom 09.09.2014.

[24] Vgl. BSI: IT-Grundschutz-Kataloge, Gefährdung G 4.90 „Ungewollte Preisgabe von Informationen durch Cloud Cartography“.

Der Betreiber des VeriMetrix-Systems könnte den Schlüsselspeicher wiederum in eine Cloud auslagern, sogar zum betreffenden Cloud-Betreiber selbst – beispielsweise in Form eines HSM-Dienstes, der nicht vom Cloud-Betreiber, sondern nur vom Kunden (oder von einem Treuhänder) konfigurierbar ist.[25]

5 Reaktion auf unrechtmäßige Standorte

Sollte das Ergebnis der Metrikenmessung Auffälligkeiten aufdecken, so muss der Cloud-Nutzer – als verantwortliche Stelle der Auftragsdatenverarbeitung – das Messergebnis juristisch interpretieren und dem konkreten Verstoß entsprechend reagieren. Hierfür sollte er zunächst den Schweregrad des festgestellten Datenschutzverstoßes anhand des vorliegenden Einzelfalls bewerten. Kriterien hierzu sind u. a. der Schutzbedarf personenbezogener Daten, der den Betroffenen drohende Schaden, die Umstände des Verstoßes und die Gesamtheit der getroffenen Sicherheitsmaßnahmen. [KSS15] Je nach Ergebnis der Schweregrad-Bewertung kommen unterschiedliche Reaktionen auf einen festgestellten Datenschutzverstoß in Betracht. Unter Berücksichtigung aller Kriterien des Einzelfalls kommen zunächst die Aufforderung zur unverzüglichen Beseitigung des Verstoßes sowie die Kontrolle seiner Beseitigung in Betracht. Bis zur Beseitigung des Verstoßes sollten nach Möglichkeit keine weiteren personenbezogenen Daten an den Cloud-Anbieter weitergegeben werden. Bei erheblichen Verstößen oder wenn der Cloud-Anbieter den Mangel nicht beseitigt, sollte der Cloud-Nutzer den Vertrag kündigen.

Zumindest bei Cloud-Speicherdiensten bestehen für den Cloud-Nutzer weitere Möglichkeiten, aktiv auf einen erheblichen Verstoß zu reagieren: Da das Herunterladen und das Löschen von zuvor gespeicherten Daten zum typischen Funktionsumfang dieser Art von Cloud-Diensten gehört, ist es dem Cloud-Nutzer möglich, die Daten beim Cloud-Anbieter ersatzlos zu löschen, auf eigene Systeme zu migrieren und danach beim Cloud-Anbieter zu löschen oder zu einem anderen Cloud-Anbieter zu migrieren.[26] [KSS15]

6 Ausblick

Wie eingangs erwähnt, sind Anwender, die personenbezogene Daten in der Cloud speichern oder verarbeiten, dazu verpflichtet, die Verfahrensabläufe und technischen Sicherheitsmaßnahmen ihres Cloud-Anbieters regelmäßig zu kontrollieren, um die Einhaltung geltender datenschutzrechtlicher Anforderungen sicherzustellen. In diesem Beitrag wurden Verfahren und eine Systemarchitektur vorgestellt, mit denen ein wichtiger Teilaspekt dieser Aufgabe automatisiert erfüllt werden kann, nämlich die Überprüfung der Zulässigkeit der Verarbeitungsstandorte der Daten. Grundsätzlich ist hierfür eine Vielzahl an Messverfahren möglich. Mithilfe praktischer Tests wird in der Folge im Rahmen des hier dargestellten Projekts VeriMetrix erprobt, welche Kombination der vorstehend skizzierten Verfahren in besonderem Maße geeignet ist, Verarbeitungsstandorte mit einem hohen Grad

[25] Derartige Funktionalität bietet auch der AWS CloudHSM-Service: `https://aws.amazon.com/de/cloudhsm`

[26] In allen drei Fällen empfehlen sich vertragliche Vereinbarungen mit dem Cloud-Anbieter über die Löschung von Backups.

an Zuverlässigkeit und Manipulationssicherheit zu bestimmen. In diesem Zusammenhang sind nicht nur die Schwellwerte zu bestimmen, ab denen Aussagen zu Standorten hinreichend plausibel sind, sondern ist auch auf eine möglichst breite und langfristige Anwendbarkeit der Messverfahren zu achten, also darauf, dass sie providerübergreifend funktionieren und flexibel an neuartige technische Gegebenheiten angepasst werden können.

Für eine umfassende Kontrolle und Bewertung der Datenschutzqualität eines Cloud-Dienstes sind weitere Indikatoren und Kennzahlen erforderlich, etwa solche zur Datentrennung, zum Zugriffsschutz oder auch zur Vollständigkeit der Datenlöschung nach Beendigung der Benutzung eines Dienstes. Auch hierfür lassen sich wichtige Kennzahlen automatisiert erheben, beispielsweise kann mithilfe von Wasserzeichen die Qualität der angewendeten Löschverfahren getestet werden. Automatisierte Kontrollen erlauben insbesondere eine zeitnahe Reaktion auf datenschutzrelevante Ereignisse bei einem Cloud-Anbieter.

Insbesondere für die Bewertung der organisatorischen und prozeduralen Sicherheitsvorkehrungen eines Cloud-Anbieters sind auch Indikatoren erforderlich, die entweder nur teilautomatisiert oder nur „händisch“ in unabhängigen Audits erhoben werden können. Automatisiert erhobene und auf anderen Wegen ermittelte Kennzahlen zusammengenommen können, bei geeigneter Vereinheitlichung und Standardisierung Grundlagen schaffen für eine höhere Transparenz und Vergleichbarkeit der Datenschutzqualität von Cloud-Anbieter und damit Hemmschwellen für die Benutzung von Cloud-Diensten beseitigen.

Literaturverzeichnis

[Bo12] Borges, G. et al.: Datenschutzrechtliche Lösungen für Cloud Computing. Kompetenzzentrum Trusted Cloud, Oct 2012.

[BS14] BSI: IT-Grundschutz-Kataloge: 14. Ergänzungslieferung Stand 2014. Bericht, Bundesamt für Sicherheit in der Informationstechnik, 2014.

[CI10] CIS: The CIS Security Metrics. The Center for Internet Security (Hrsg.), 2010.

[Fl12] Flach, P.: Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press, 2012.

[IS09] ISO: ISO 27004:2009, Information Security Management – Measurement. International Organization of Standardization (Hrsg.), 2009.

[Ja07] Jaquith, A.: Security Metrics: Replacing Fear, Uncertainty, and Doubt. 2007.

[JSW15] Jäger, B.; Selzer, A.; Waldmann, U.: Die automatisierte Messung von Cloud-Verarbeitungsstandorten. Datenschutz und Datensicherheit – DuD, 01:26–30, 2015.

[KNW13] Kunz, T.; Niehues, P.; Waldmann, U.: Technische Unterstützung von Audits bei Cloud-Betreibern. Datenschutz und Datensicherheit – DuD, 37(8):521–525, 2013.

[KSS15] Kunz, T.; Selzer, A.; Steiner, S.: Konsequenzen festgestellter Verstöße bei der Auftragsdatenverarbeitungskontrolle. Datenschutz und Datensicherheit – DuD, 01:21–25, 2015.

[LH15] Luhn, S.; Hils, M.: Ein Rahmenwerk für Datenschutz-Metriken in der Cloud. In (Douglas Cunningham, Petra Hofstedt, Klaus Meer Ingo Schmitt, Hrsg.): INFORMATIK 2015, Lecture Notes in Informatics (LNI). Gesellschaft für Informatik, 2015.

[Ma11] Massonet, P.; Naqvi, S.; Ponsard, C.; Latanicki, J.; Rochwerger, B.; Villari, M.: A Monitoring and Audit Logging Architecture for Data Location Compliance in Federated Cloud Infrastructures. In: Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), 2011 IEEE International Symposium on. S. 1510–1517, may 2011.

[NI08] NIST: NIST 800-55: Performance Measurement Guide for Information Security. National Institute of Standards and Technology (Hrsg.), 2008.

[Ri09] Ristenpart, T.; Tromer, E.; Shacham, H.; Savage, S.: Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In: Proceedings of the 16th ACM conference on Computer and communications security. CCS '09, ACM, New York, NY, USA, S. 199–212, 2009.

[Ri11] Ries, T. et al.: Verification of data location in cloud networking. In: Fourth IEEE International Conference on Utility and Cloud Computing, S. 439–444, 2011.

[Se13] Selzer, A.: Die Kontrollpflicht nach §11 Abs. 2 Satz 4 BDSG im Zeitalter des Cloud Computing – Alternativen zur Vor-Ort-Kontrolle des Auftragnehmers durch den Auftraggeber. Datenschutz und Datensicherheit – DuD, 4:215–219, 2013.

[Se14] Selzer, A.: Datenschutz bei internationalen Cloud Computing Services. Datenschutz und Datensicherheit – DuD, 38(7):470–474, 2014.

[SK99] Schneier, B.; Kelsey, J.: Secure audit logs to support computer forensics. ACM Transactions on Information and System Security (TISSEC), 2(2):159–176, 1999.

[So11] Sowa, A.: Metriken, der Schlüssel zum erfolgreichen Security und Compliance Monitoring. 2011.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Dynamische Zertifizierung von Cloud Computing-Diensten: Eine rechtswissenschaftliche Betrachtung am Beispiel „Verfügbarkeit“

Ass. jur. Johanna M. Hofmann LL.M.[1]

Abstract: Dieser Beitrag betrachtet die dynamische Zertifizierung von Cloud-Diensten am Beispiel des Kriteriums „Verfügbarkeit“ aus rechtswissenschaftlicher Sicht.[2] Nach der Einführung in die Problematik der Zertifizierung von Cloud Computing-Diensten (2) und der Vorstellung der Grundzüge der dynamischen Zertifizierung (3) wird das methodische Vorgehen zur rechtlichen Überprüfung von automatisiert und kontinuierlich zu prüfenden Anforderungen dargestellt (4). Diese Methodik wird sodann auf das Beispiel „Verfügbarkeit“ angewendet (5).

Keywords: Dynamische Zertifizierung; Cloud Computing; Kontrollpflicht des Auftraggebers; Methode KORA; Verfügbarkeit;

1 Einleitung

Mit Cloud Computing bezeichnet man die Auslagerung der eigenen Datenverarbeitung in eine vernetzte Rechnerarchitektur, das heißt „in die Wolke“.[3] Datenschutzrechtlich handelt es sich beim Cloud Computing in aller Regel um eine sogenannte Auftragsdatenverarbeitung.[4] Vielfach sind die „in die Cloud“ übertragenen Daten personenbezogen.[5] Die Übertragung ist dann nach dem Bundesdatenschutzgesetz zu beurteilen.[6] Die Auftragsdatenverarbeitung ist ein Sonderfall der Verantwortungszuweisung. Bei Vorliegen der Voraussetzungen[7] liegt keine Übermittlung personenbezogener Daten vor. Für die Zulässigkeit der Datenverarbeitung durch den Cloud-Dienst-Anbieter[8] ist entgegen des Grundsatzes aus § 4 Abs. 1 BDSG keine gesetzliche oder rechtsgeschäftliche Ermächtigung oder gesetzliche Gebotsnorm

[1] Wissenschaftliche Mitarbeiterin in der Projektgruppe für verfassungsverträgliche Technikgestaltung (provet) im Wissenschaftlichen Zentrum für Informationstechnik-Gestaltung (ITeG) an der Universität Kassel, Pfannkuchstr. 1, 34109 Kassel, j.hofmann@uni-kassel.de.

[2] Gegenstand dieses Beitrags sind Ergebnisse des Forschungsprojekts „Next Generation Certification“ (NGCert), das vom Bundesministerium für Bildung und Forschung als Teil des Themenfeldes “Forschung für Sicheres Cloud Computing” im Rahmen der Hightech-Strategie der Bundesregierung gefördert wird (ngcert.de).

[3] *Weichert*, DuD 2010, 679. Zur gängigen Definition, den drei Dienstmodellen (Software as a Service, Infrastructure as a Service und Plattform as a Service) und den vier Erscheinungsformen (Private-/Public-/Hybrid- und Community Cloud) *Mell/Grance* (National Institute of Standards and Technologie (NIST), The NIST Definition of Cloud Computing, September 2011, 2 ff..

[4] § 11 Bundesdatenschutzgesetz (BDSG).

[5] Vgl. § 3 Abs. 1 BDSG.

[6] *Roßnagel*, Handbuch Datenschutzrecht, 2003, Kap. 7.9, Rn. 59.

[7] §§ 3 Abs. 8 Satz 3, 11 BDSG.

[8] „Cloud-Anbieter“.

erforderlich. Übertragen auf das Cloud Computing bedeutet dies, dass der Cloud-Dienst-Kunde[9] (Auftraggeber) gemäß § 11 Abs. 1 Satz 1 BDSG die datenschutzrechtliche Verantwortlichkeit über diejenigen personenbezogenen Daten behält, die er oder die Anwender „in die Cloud" übertragen. Er bleibt für diese Daten derjenige, der über Zweck und Mittel der Datenverarbeitung entscheidet.[10] Der Cloud-Anbieter (Auftragnehmer) hingegen ist in diesem Verhältnis grundsätzlich keine verantwortliche Stelle, sondern hat die Daten entsprechend dem Vertrag über die Auftragsdatenverarbeitung sowie den Weisungen des Cloud-Kunden zu verarbeiten. Davon zu unterscheiden sind solche Daten, die nicht „in die Cloud" übertragen werden, sondern durch die Nutzung des Dienstes entstehen (z.B. Nutzungsdaten). Bei diesen kann es sich ebenfalls um personenbezogene Daten handeln.[11] Für deren Verarbeitung wird in der Regel der Cloud-Anbieter verantwortliche Stelle sein.[12]

Der Auftraggeber ist gemäß § 11 Abs. 2 Satz 1 BDSG verpflichtet, den Auftragnehmer „sorgfältig auszuwählen". Der Cloud-Kunde hat dabei zu überprüfen, ob der Cloud-Anbieter ein angemessenes Datenschutzniveau bietet.[13] Darüber hinaus hat sich der Auftraggeber „vor Beginn der Datenverarbeitung und sodann regelmäßig von der Einhaltung der beim Auftragnehmer getroffenen technischen und organisatorischen Maßnahmen zu überzeugen".[14] Folglich ist der Cloud-Kunde verpflichtet, vor Beginn der Auftragsdatenverarbeitung und sodann in regelmäßigen Abständen die Einhaltung der Anforderungen und seiner Weisungen seitens des Cloud-Anbieters zu kontrollieren. Wie diese Kontrolle zu erfolgen hat, ist im Gesetz nicht ausdrücklich geregelt. Die Kontrollmöglichkeiten reichen von der persönlichen Vor-Ort-Kontrolle des Cloud-Kunden, über Selbstauditierungen des Cloud-Anbieters bis hin zu Bewertungen durch Dritte. Die persönliche Vor-Ort-Kontrolle ist dabei für beide Seiten mit Nachteilen verbunden. Der Cloud-Anbieter ist dazu verpflichtet, jedem einzelnen seiner Kunden die Ausübung der gesetzlichen Kontrollpflicht zu ermöglichen und hat gegebenenfalls Zutritt zu seinen Serverräumen zu gewähren. Dies kann einen erheblichen Zeit- und Geldaufwand bedeuten. Wesentlich schwerer wiegen allerdings die damit verbundenen Sicherheitsrisiken, die der Zutritt Betriebsfremder zu seinen Betriebsräumlichkeiten mit sich bringen kann. Hinsichtlich des Cloud-Kunden stößt eine Vor-Ort-Kontrolle auf die folgenden Schwierigkeiten: Zum einen kennt er den Speicherort der Daten regelmäßig

[9] Derjenige, der in vertraglicher Beziehung zum Cloud-Anbieter steht und die Cloud-Dienste bezieht („Cloud-Kunde"). Der Cloud-Kunde kann seinerseits Kunden haben, die hier der Klarheit halber als „Anwender" bezeichnet werden. Der Cloud-Anbieter ist derjenige, der die Cloud-Dienste entweder selbst oder durch die Hinzuziehung dritter Subunternehmer erbringt. Mit anderen Worten stellt der Cloud-Anbieter Software, Infrastruktur oder die Plattform zur Verfügung.

[10] *Dammann,* in: Simitis, BDSG 8. Aufl. 2014, § 3 BDSG Rn. 224. Ausführlich zur Definition der „verantwortlichen Stelle" die Artikel 29-Arbeitsgruppe Stellungnahme 05/2012 zum Cloud Computing („WP 196"), Kap. 3, 10 ff.

[11] *Boos/Kroschwald/Wicker*, ZD 2013, 207.

[12] Hierzu Arbeitskreise Technik und Medien der Konferenz der Datenschutzbeauftragten des Bundes und der Länder sowie der Arbeitsgruppe Internationaler Datenverkehr des Düsseldorfer Kreises, Orientierungshilfe – Cloud Computing, Version 2.0, Stand 09.10.2014 („Düsseldorfer Kreis, Orientierungshilfe"), 9.

[13] *Petri,* in: Simitis, BDSG 8. Auflage 2014, § 11 BDSG Rn. 55.

[14] § 11 Abs. 2 Satz 4 BDSG.

nicht; zum anderen liegt der Speicherort möglicherweise in großer geographischer Entfernung zu seinem eigenen Standort. Es ist zudem denkbar, dass die jeweiligen Daten auf verschiedene Serverstandorte verteilt sind, es daher mit der Kontrolle eines einzigen Rechenzentrums zur Erfüllung der gesetzlichen Kontrollpflicht nicht getan wäre.

Die Unterstützung der Kontrollpflicht durch Dritte, stellt sich demgegenüber als besonders vorteilhaft dar. Auch ist die Aussage eines Dritten objektiver und daher glaubhafter als eine interessengeleitete Selbsteinstschätzung des Cloud-Anbieters.[15]

Konformitätsbewertungen im Bereich der Informationstechnologien (IT) werden in der Rechtswissenschaft uneinheitlich als Gütesiegel, Testate oder Zertifikate bezeichnet. Eine klare Abgrenzung ist in Ermangelung einer einheitlichen gesetzlichen Regelung noch nicht erfolgt. Auch unterscheiden sich tatbestandliche Voraussetzungen, Verfahren und Rechtsfolgen. Dieser Beitrag wird die Begriffe synonym für Bestätigungen über das Erfüllt-Sein von festgelegten Kriterien verwenden, die im Rahmen eines geregelten Konformitätsbewertungsverfahren durch eine akkreditierte Stelle vergeben werden.

2 Die Zertifizierung von Cloud Computing-Diensten

Bereits in den Gesetzesmaterialien,[16] aber auch unter den Aufsichtsbehörden[17] und in der Literatur[18] ist anerkannt, dass im Einzelfall alternative Gestaltungsformen zur persönlichen Vor-Ort-Kontrolle des Cloud-Kunden zulässig sein können. Die Heranziehung fachkundiger Aussagen Dritter, in Form von Zertifizierungen, ist eine dieser Möglichkeiten. Diese Möglichkeit scheint besonders vorteilhaft, da sie neben einer bloßen Erleichterung im Vergleich zur persönlichen Vor-Ort-Kontrolle mit der Aussage eines fachkundigen Dritten über die Normkonformität eines Anbieters, eines Systems oder eines Dienstes einen Vertrauenstatbestand schaffen[19] und über ein Informationsdefizit bei Cloud-Kunden und Anwendern hinweghelfen soll. Dem Cloud-Kunden wird es regelmäßig an Fachkunde und Einblicken in die Unternehmensorganisation des Cloud-Anbieters mangeln. Das Zertifikat eines fachkundigen Dritten kann ihm daher mehr Sicherheit bieten, als er durch seine persönliche Vor-Ort-Kontrolle erreichen könnte.

Allerdings verlangt ein Zertifikat dem Cloud-Kunden einen Vertrauensvorschuss ab. So soll der Cloud-Kunde darauf vertrauen, dass die Aussage eines Zertifikats auch noch zu

[15] Eine Selbsteinschätzung des Cloud-Anbieters wird im Zusammenhang mit den Kontrollpflichten des Cloud-Kunden völlig zu Recht ohnehin weitgehend als unzureichend erachtet. Von einer Diskussion und Entscheidung dieser Frage wird hier aus Gründen des Platz- und Relevanzmangels Abstand genommen.
[16] BT-Drs. 16/13657, 28. So auch Art. 26 Abs. 2 aa des Entwurfs zur europäischen Datenschutzgrundverordnung (konsolidierte Fassung des Rates vom 19. Dezember 2014, „DSGVO-E“).
[17] Düsseldorfer Kreis, Orientierungshilfe, 10. Dabei betont der Arbeitskreis, dass das alleinige Vorliegen von Zertifikaten nicht von der Kontrollpflicht entbindet.
[18] So etwa WP 196, Kap. 4.2, 27.
[19] OLG Dresden, MMR 2012, 6780.

einem späteren Zeitpunkt (etwa ein Jahr später) immer noch zutrifft.[20] Tatsächlich steht die dynamische Entwicklung auf dem Gebiet der IT dabei jedoch oft im Widerspruch zur statischen Aussage einer derartigen Konformitätsbewertung. Während sich IT und Rechtswissenschaft in Wirklichkeit unaufhaltsam weiterentwickeln und sich die tatsächlichen Gegebenheiten verändern können, wird ein klassisches Zertifikat infolge einer Überprüfung in der Regel für einen festgelegten Zeitraum erteilt, z.B. zwei Jahre, und bestätigt einen möglicherweise längst vergangenen und überholten Zustand. Das Ergebnis – Unsicherheit – steht dem eigentlichen Zweck von Zertifizierungen, nämlich der Steigerung von Markvertrauen in Dienste und Dienstleister – entgegen. Unsicherheit ist eine Eigenschaft, die gerade nicht mit Zertifizierungen in Verbindung gebracht werden sollte und letzten Endes dazu führt, dass das Instrument der Zertifizierung an Reputation und Wirksamkeit verliert.[21]

3 Die dynamische Zertifizierung

Eine dynamische Zertifizierung ist ein möglicher Lösungsansatz. Sie zeichnet sich durch die kontinuierliche und automatisierte Überprüfung einzelner, zuvor festgelegter Anforderungen – insbesondere, aber nicht ausschließlich aus Datenschutz und Datensicherheit – aus. Zudem schafft sie Marktanreize[22] und zielt darauf ab, dem Cloud-Kunden die Ausübung seiner Kontrollpflicht aus § 11 Abs. 2 Satz 4 BDSG zu erleichtern.

Ein dynamisches Zertifizierungssystem erfordert ein idealerweise kontinuierliches Überwachen (Monitoring) der überprüften Systeme und Dienste mittels Messtechniken und Berechnungsmethoden (Metriken).[23] Anhand der Messergebnisse muss feststellbar sein, ob ein für die Zertifikatserteilung relevanter vorgegebener Sollwert erfüllt wird. Entscheidend ist, dass die Dynamik eines Cloud-Dienstes berücksichtigt wird. Es müssen folglich unter anderem Veränderungen rechtlicher, technischer und tatsächlicher Art bedacht werden. Das wiederum setzt voraus, dass sich die zu prüfenden Kriterien, Messergebnisse und Messtechniken während der Messung verändern können und diese Veränderungen Berücksichtigung finden. Wird ein Merkmal A zu einem Zeitpunkt X überprüft, liegt beispielsweise Dynamik vor, wenn die Prüfung des Merkmals A zu einem Zeitpunkt X+1 (der zeitlich nach X liegt) ein unterschiedliches Ergebnis erzeugt. Es können sich dabei verschiedene Konstellationen ergeben: Eine Metrik kann dazu dienen, mehrere Kriterien zu überprüfen. Ein Kriterium kann aber auch mit nur einer einzigen Metrik überprüfbar sein. Zur Darstellung desselben (rechtlichen) Kriteriums können verschiedene Metriken anwendbar sein. Und schließlich kann ein Kriterium aber

[20] *Dahm,* DuD 2002, 413 spricht insoweit von Vertrauen auf „Quasi-Kulanz" – einem (vor)vertragliches Verhalten desjenigen, der das Zertifikat verwendet.
[21] Auch nach Art. 39 Abs. 4 DSGVO-E ist die Zertifizierung bei Wegfall der Voraussetzungen zu entziehen.
[22] Hierzu *Hornung,* ZD 2014, 219 ff.
[23] Ausführlich hierzu *Stephanow/Gall*, Language Classes for Cloud Service Certification Systems, 2015 IEEE 11th World Congress on Services (SERVICES), i.E.

auch dadurch erfüllt sein, dass ein oder mehrere andere Kriterien alleine oder in Kombination anhand von einer oder mehreren Metriken alleine oder in Kombination nachgewiesen werden.

Das Vertrauen des Marktteilnehmers ist berechtigt, sobald die durch ein Zertifikat bestätigten, festgelegten Kriterien entweder innerhalb möglichst klein zu definierender Intervalle oder zu einem beliebigen Zeitpunkt bedarfsabhängig überprüfbar werden. Es wird ihm nicht abverlangt, blind auf die Zertifikatsaussage zu vertrauen, sondern ein Mittel zur Überwindung seines Informationsdefizits in die Hand gegeben. Die Werbung mit einem dynamischen Zertifikat ist damit berechtigt und lauter. Die dynamische Zertifizierung schafft folglich für die jeweiligen Teilbereiche Transparenz und Rechtssicherheit auf dem bislang undurchsichtigen Gebiet des Cloud Computing. Ziel eines dynamischen Zertifizierungskonzepts muss es demnach sein, den Prüfkatalog stets aktuell zu halten und – der Dynamik entsprechend – zu erweitern.

4 Die Methode KORA

Die Entwicklung eines dynamischen Zertifizierungssystems erfordert die Erstellung eines umfassenden Kriterienkatalogs. Hierzu bietet sich die Verwendung der Methode zur Konkretisierung rechtlicher Anforderungen zu technischen Gestaltungsvorschlägen (KORA)[24] an. Dahinter steht die Überlegung, dass rechtliche Vorgaben in aller Regel zu abstrakt sind, um konkrete technische Gestaltungsmerkmale für ein konkretes technisches System zu enthalten.

Angesichts der gegebenenfalls unterschiedlichen Auswirkungen ist strikt zwischen Prüfungsinhalt (etwa Diensten und Systemen) und Zertifizierungsverfahren zu differenzieren.[25]

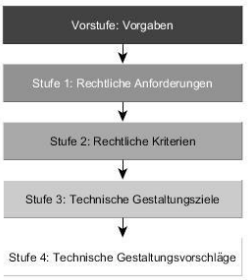


Abb. 1: Die Methode KORA

[24] Die Methode KORA wurde von der Projektgruppe verfassungsverträgliche Technikgestaltung an der Universität Kassel entwickelt. Zur Methode siehe *Hammer/Pordesch/Roßnagel*, Informatik und Gesellschaft, 1993, 21 ff.; *dies.*, Betriebliche Telefon- und ISDN-Anlagen rechtsgemäß gestalten, 1993, 46 ff.

[25] Dies gründet darin, dass für beide unterschiedliche rechtliche Grundlagen gelten können.

Abbildung 1 zeigt die Methode KORA, bei der in einem vierstufigen Prüfungsaufbau rechtliche Anforderungen aus der Verfassung abgeleitet und über zwei weitere Stufen schließlich zu konkreten technischen Gestaltungsvorschlägen konkretisiert werden. Diese Vorschläge müssen zudem dynamische Aspekte betreffen,[26] kontinuierlich überwachbar[27] und schließlich automatisiert überprüfbar[28] sein.

KORA verfolgt das Ziel rechtsverträglicher und nicht „nur" rechtmäßiger Gestaltung. Rechtmäßigkeit beschreibt zwingend zu erfüllende Voraussetzungen („Muss-"Regelungen). Daneben sind „Soll-"Anforderungen solche, die in der Regel vorliegen müssen, von denen in Ausnahmefällen aber abgewichen werden kann. Rechtsverträglichkeit beschreibt einen Zustand, der erstrebenswert ist. Insoweit ist auch auf „Kann-"Regelungen zurückzugreifen, die nicht zwingend sind. Das deutsche Datenschutzrecht sieht zwingende Grundsätze vor, die einzuhalten sind. Andere Aspekte wiederum sind lediglich vorteilhaft. Auf KORA übertragen bedeutet dies, dass ein Großteil der rechtlichen Aspekte zwingend einzuhalten sein wird. Darüber hinaus sind aber auch Gestaltungsformen denkbar, die gegenüber anderen ein Mehr an Datenschutz oder -sicherheit bieten und die deshalb vorzugswürdig erscheinen. Der große Vorteil dieses methodischen Ansatzes liegt darin, dass sie dem Juristen, der normalerweise erst zu Rate gezogen wird, wenn Technik erstellt und implementiert wurde, ermöglicht, zu einem möglichst frühen Zeitpunkt gestaltend mitzuarbeiten. Ziel von KORA ist in diesem Fall, Gestaltungsvorschläge für eine datenschutzfreundliche Technik zu gewinnen.

Rechtliche Kriterien (Stufe 2) leiten sich bei KORA aus rechtlichen Anforderungen ab (Stufe 1), die wiederum direkt verfassungsrechtlichen Vorgaben (Vorstufe) entstammen, und beziehen sich einerseits auf Technik, andererseits auf soziale und rechtliche Aspekte, indem sie auf abstrakter Ebene Lösungswege für die Anforderungen aufzeigen, ohne jedoch konkrete Lösungsansätze zu benennen. Rechtlichen Kriterien kommt eine Doppelfunktion dahingehend zu, dass sie zum einen nachfolgende Konkretisierungen bewerten – und damit ein gewisses Maß an Konkretisierung erfahren haben, zum anderen liefern sie Hinweise darauf, wie eine gute Lösung aussehen muss.[29]

Technische Gestaltungsziele (Stufe 3) sind abstrakte technische Anforderungen und stellen einen Zwischenschritt zur Systematisierung technischer Gestaltung dar. Bezugspunkte von technischen Gestaltungszielen sind grundlegende Systemfunktionen, die Systemarchitektur oder aber Daten.[30] Nur selten handelt es sich dabei um zwingend einzuhaltende Aspekte.

Technische Gestaltungsempfehlungen (Stufe 4) sind das Ergebnis der Bewertung technischer Merkmale anhand der gefundenen technischen Gestaltungsziele und

[26] Andernfalls läge mangels Veränderbarkeit kein Überprüfungsbedürfnis vor.
[27] Das wäre beispielsweise nicht der Fall, wenn ein Aspekt zu einem späteren Zeitpunkt nicht mehr existiert.
[28] Gemeint ist hier ein vollständiger Automatisierungsgrad, bei dem keine menschliche Intervention nötig ist.
[29] *Hammer/Pordesch/Roßnagel*, Informatik und Gesellschaft, 1993, 21.
[30] *Pordesch,* Die elektronische Form und das Präsentationsproblem, Nomos, Baden-Baden 2003, 266.

gleichzeitig die letzte Prüfungsstufe von KORA. Angesichts der Fülle an technischen Lösungsansätzen und deren Kombinationsmöglichkeiten ist eine abschließende Betrachtung nicht möglich. Eine Sammlung von Gestaltungsvorschlägen kann demnach lediglich eine stark begrenzte Auswahl darstellen und erhebt keinen Anspruch auf Vollständigkeit.

5 KORA am Beispiel „Verfügbarkeit“

Verfügbarkeit ist ein rechtliches Kriterium (Stufe 2). Es folgt für den Prüfungsinhalt aus den Anforderungen Datenschutz und Geheimnisschutz (Stufe 1), die wiederum auf den Grundrechten auf informationelle Selbstbestimmung beziehungsweise auf Eigentum und auf Berufsfreiheit (Vorstufe) basieren. Verfügbarkeit als eines der IT-Sicherheitsziele[31] gehört sinnvollerweise zu einem umfassenden Datenschutzsystem. Es soll Informationen vor Verlust, Entzug, Blockade und Zerstörung schützen und gleichzeitig (in seiner zweiten Ausprägung) dafür sorgen, dass ein System bei autorisiertem Zugriff innerhalb einer bestimmten Zeit zur Verfügung steht und technisch vor der Beeinträchtigung der Verfügbarkeit durch Nichtberechtigte geschützt ist. Dem Cloud-Kunden müssen folglich die Cloud-Dienste sowie die Daten zu jeder Zeit in dem vertraglich vereinbarten Umfang zur Nutzung zur Verfügung stehen. Obgleich es sich bei der Verfügbarkeit um ein IT-Sicherheitsziel handelt, richten sich die Voraussetzungen und die Rechtsfolgen nach vertraglich zwischen dem Cloud-Anbieter und dem Cloud-Kunden zu vereinbarenden Erfüllungsgraden (Service Levels). Die Möglichkeit des Zugriffs auf Cloud-Ressourcen über ein öffentliches Netzwerk, wie etwa bei der Public Cloud, und die damit verbundenen erhöhten Gefahren stellt den Cloud-Anbieter in diesem Zusammenhang vor eine große Herausforderung. Fehlerhafte Systemkonfigurationen und eine große Anzahl an Cloud-Serviceanfragen können der Verfügbarkeit (sowohl der Daten als auch der Systeme) abträglich sein. Strafrechtlich kann ein Angriff auf die Verfügbarkeit beispielsweise über § 265a StGB (Erschleichung von Leistungen) oder § 303a StGB (Datenveränderung) von Bedeutung sein.

„Verfügbarkeit“ ist für das Cloud Computing (den Prüfungsinhalt) von hervorgehobener Bedeutung. Es tritt in zwei verschiedenen Formen auf. Zum einen bezieht es sich auf Daten, zum anderen auf Systeme.[32] Hier wird das Kriterium „Verfügbarkeit“ in Bezug auf Daten besprochen. Dabei kann es vorkommen, dass Gestaltungsziele (Stufe 3) oder -vorschläge (Stufe 4) auf beide Erscheinungsformen anwendbar sind.[33]

Die Anwendung von KORA stellt sich für das hier darzustellende Beispiel „Verfügbarkeit“ bezüglich des zu zertifizierenden Inhalts (Cloud-Dienst) wie folgt dar.

[31] § 2 Abs. 2 Gesetz über das Bundesamt für Sicherheit in der Informationstechnik.
[32] Wie es Gegenstand des Entwurfs des IT-Sicherheitsgesetzes (BT-Drks. 18/4096) ist.
[33] So schützt beispielsweise die Ausfallsicherheit in erster Linie die Verfügbarkeit von Systemen, ohne letztere, wären allerdings auch keine Daten verfügbar, so dass das Gestaltungsziel auch für die Verfügbarkeit von Daten relevant ist.

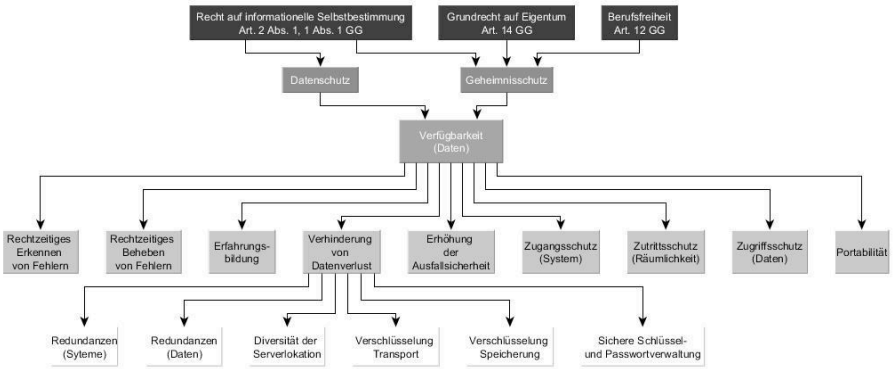


Abb. 2: Die Methode KORA am Beispiel „Verfügbarkeit“ für den Prüfungsinhalt

Hinweis: Abbildung 2 zeigt die Anwendung der Methode KORA auf das Beispiel der Verfügbarkeit (Stufe 2) bezogen auf den Prüfungsinhalt. Während die verfassungsrechtlichen Grundlagen (Vorstufe) sowie die zugrundeliegenden rechtlichen Anforderungen (Stufe 1) abschließend sind, kann dies für die aus dem Kriterium der Verfügbarkeit abzuleitenden technischen Gestaltungsziele (Stufe 3) und Gestaltungsvorschläge (Stufe 4) nicht gelten. Vielmehr wurde die Konkretisierung des Kriteriums „Verfügbarkeit“ aus Gründen der Übersichtlichkeit auf neun Gestaltungsziele beschränkt. Davon wurde – wiederum exemplarisch – lediglich jenes der „Verhinderung von Datenverlust“ zu sechs (nicht abschließenden) Gestaltungsvorschlägen konkretisiert.

Aus dem rechtlichen Kriterium der Verfügbarkeit (gegebenenfalls in Verbindung mit zusätzlichen Kriterien, wie etwa denjenigen der Integrität der Daten und Systeme, der Transparenz, der Überprüfbarkeit, der Beherrschbarkeit und Steuerungsfähigkeit, der Beweissicherung) lassen sich unter anderem die technischen Gestaltungsziele „rechtzeitiges Erkennen von Fehlern“, „rechtzeitiges Beheben von Fehlern“, „Erfahrungsbildung“, „Verhinderung von Datenverlust“, „Erhöhung der Ausfallsicherheit“, „Zugangsschutz“, „Zutrittsschutz“, „Zugriffsschutz“ und „Portabilität“ ableiten.

Während sich die ersten beiden Ziele von selbst erklären, ist mit „Erfahrungsbildung“ gemeint, dass beispielsweise gezielte Angriffe auf Systeme einen Lerneffekt erzielen, der Verfügbarkeit unterstützen kann. Zur „Verhinderung von Datenverlust“ sogleich. „Erhöhung der Ausfallsicherheit“ kann beispielsweise durch physische Serversicherungsmaßnahmen erfolgen. „Zugangsschutz“ bezieht sich auf Systeme, „Zutrittsschutz“ auf Räumlichkeiten und „Zugriffsschutz“ auf Daten. Gemeint ist jeweils Schutz vor unberechtigtem Zugang, Zutritt und Zugriff. „Portabilität“ betrifft Daten und die Möglichkeit deren Verlagerung von einem Cloud-Anbieter zu einem anderen. Offene Schnittstellen und Datenformate können einen sog. Vendor-lock in verhindern, der dadurch entsteht, dass ein Cloud-Anbieter beispielsweise unübliche Formate verwendet, die nicht mit den Systemen anderer Anbieter kompatibel sind und damit verhindert oder

erschwert, dass ein Cloud-Kunde zu einem anderen Anbieter wechselt.

Das Gestaltungsziel „Verhinderung von Datenverlust“ wurde exemplarisch weiter zu sechs nicht abschließenden Gestaltungsvorschlägen konkretisiert. Dies sind namentlich „Redundanzen“ – jeweils von Systemen und Daten, „Diversität der Serverlokation“, „Verschlüsselung“ – jeweils auf dem Transport und während der Speicherung und die „sichere Schlüssel- und Passwortverwaltung“.

Verfügbarkeit sollte dadurch gefördert werden, dass Systeme und Daten jeweils redundant vorgehalten werden. Es werden dabei von Daten beispielsweise Kopien angefertigt, die verhindern sollen, dass die Daten durch einen Vorfall unwiederbringlich gelöscht werden. Verteilen sich die Systeme und die darin gespeicherten Daten zudem auf verschiedene Serverstandorte, ist gleichzeitig die Gefahr verringert, die beispielsweise von Naturkatastrophen für die Daten und Systeme ausgeht („Diversität der Serverlokation“). Daten müssen zudem auf dem Transport ebenso wie bei deren Speicherung „in der Cloud“ dem jeweiligen Stand der Technik entsprechend verschlüsselt werden. Schließlich sind die Schlüssel und Passwörter sicher zu verwalten.

Wie bereits erwähnt, handelt es sich hierbei um eine nicht abschließende, exemplarische Ableitung. Ziel war es, das methodische Vorgehen zu veranschaulichen. Hinsichtlich des analysierten Beispiels „Verfügbarkeit“ in Bezug auf den Prüfungsinhalt unterscheidet sich die dynamische freilich nicht von der statischen Zertifizierung. Beide haben die Bestätigung der Normkonformität von Cloud-Diensten zum Gegenstand. Die Betrachtung von „Verfügbarkeit“ im Zusammenhang mit dem dynamischen Verfahren wird zwar in weiten Teilen mit den hier dargestellten Ableitungen zum Inhalt übereinstimmen, darüber hinaus aber auch Unterschiede aufweisen. Diese Unterschiede rühren daher, dass nach bisherigem Erkenntnisstand die Daten, deren Übermittlung an den Zertifizierer zum Zwecke der dynamischen Zertifizierung erforderlich ist, keine personenbezogenen Daten im Sinne des Bundesdatenschutzgesetzes sind. Solche sind aber erforderlich damit letzteres Anwendung findet. Gemäß § 3 Abs. 1 BDSG sind personenbezogene Daten „Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener).“ Die im Rahmen eines dynamischen Zertifizierungsverfahrens zu übermittelnden Daten beziehen sich jedoch aller Voraussicht nach nicht auf natürliche Personen. Es handelt sich vielmehr um Prozess- und Organisationsdaten aus dem Unternehmen des Cloud-Anbieters. Jedenfalls weisen sie für den Zertifizierer keinen Personenbezug auf. Selbst für den Fall, dass die Übermittlung von Daten, die aus Sicht des Zertifizierers den Rückschluss auf bestimmte oder bestimmbare Personen (etwa einen Mitarbeiter des Cloud-Anbieters) zulassen, nicht ausgeschlossen werden kann, wird Kenntnis des Zertifizierers dieser Daten voraussichtlich nicht erforderlich sein. Vielmehr sind diese Daten vor ihrer Übermittlung an den Zertifizierer zu anonymisieren. Demzufolge ist das Bundesdatenschutzgesetz insoweit nicht anwendbar. Daraus folgt für das rechtliche Kriterium der Verfügbarkeit, dass es sich für das dynamische Zertifizierungsverfahren aus der Anforderung „Datenschutz“ nur indirekt und zwar insoweit ableitet, dass der zu zertifizierende Cloud-Dienst selbst datenschutzgerecht ausgestaltet sein muss. Hingegen folgt Verfügbarkeit in

Bezug auf das Verfahren vorrangig aus den Anforderungen „Geheimnisschutz“, „Nachvollziehbarkeit und Kontrollierbarkeit der Zertifizierung“ und „unternehmerische Entscheidungsfreiheit“. Der Aspekt der Dynamik spiegelt sich daneben in den Kriterien „Überprüfbarkeit“,[34] „Veränderbarkeit“,[35] „Adäquanz“[36] und „Aktualität“[37] wider, die sich wiederum aus den Anforderungen „Rechtssicherheit“, „Nachvollziehbarkeit“ und „Kontrollierbarkeit der Zertifizierung“ sowie „Verbraucherschutz“ ableiten. Das dynamische Zertifizierungsverfahren muss hinsichtlich Aufbau und Organisation, Rechtsfolgen und Voraussetzungen, Kompetenzen und Haftung transparent sein. Die Prüfungsergebnisse müssen verfügbar sein – für welchen Adressatenkreis, bleibt zu untersuchen. Klar ist indes, dass ein dynamisches Zertifizierungsverfahren, dessen Ergebnisse nicht jederzeit abrufbar sind, nicht den vollständigen Mehrwert gegenüber einem statischen Verfahren entfalten kann. Denn nur durch Verfügbarkeit der Prüfungsergebnisse und der zugrundeliegenden Richtlinien kann Vertrauen in das dynamische Zertifikat und damit in den zertifizierten Dienst gewonnen werden. Anwender und Cloud-Kunde müssen in Erfahrung bringen können, was die dynamische Prüfung ergeben und anhand welcher Regeln sie stattgefunden hat. Gleiches gilt für die Verfügbarkeit von Systemen. Nur wenn die für die dynamische Zertifizierung erforderlichen Systeme verfügbar sind, kann eine automatisierte dynamische Überprüfung überhaupt stattfinden.

Auf der Ebene der technischen Gestaltungsziele erlangen von den oben für den Inhalt abgeleiteten Zielen insbesondere all diejenigen besondere Bedeutung für das Zertifizierungsverfahren, die Aspekte der Dynamik und des Lernprozesses widerspiegeln. Das sind insbesondere die rechtzeitige Fehlererkennung und -behebung sowie die Erfahrungsbildung.

Ein dynamisches Zertifizierungsverfahren erfordert eine abschließende Betrachtung und Beurteilung aller rechtlichen Vorgaben. In einem weiteren Schritt sind die einzelnen Vorschläge im interdisziplinären Diskurs auf ihre Wirtschaftlichkeit und Funktionalität hin zu untersuchen. Nur so kann überhaupt ein rechtmäßiges, wirtschaftliches und technisch umsetzbares System entstehen, das sich von vorhandenen Verfahren absetzt.

[34] Die Durchführung einer Überprüfungsmaßnahme erfordert Überprüfbarkeit. Der Einzelne muss die Prüfungsschritte nachvollziehen können, um darüber zu entscheiden, ob er dagegen vorgehen will oder nicht.

[35] Hier wird der Unterschied der dynamischen zur statischen Zertifizierung besonders deutlich. Das Kriterium beinhaltet Dynamik und bezieht sich gleichfalls auf rechtliche, technische und tatsächliche Gegebenheiten. Das dynamische Zertifizierungsverfahren muss anpassbar, das heißt grundsätzlich offen für Neuerungen sein.

[36] „Adäquanz“ wird hier im Sinne einer generellen Geeignetheit begriffen. Die einzelnen Komponenten des dynamischen Zertifizierungsverfahrens müssen so ausgewählt sein, dass eine Förderung der angestrebten dynamischen Gestaltung dadurch nicht außerhalb aller Wahrscheinlichkeit liegt. „Veränderbarkeit“ ist dabei notwendige Bedingung für „Adäquanz“. Im Unterschied zu ersterer, geht es bei Adäquanz statt um die bloße Möglichkeit der Berücksichtigung der Dynamik um eine entsprechende Wahrscheinlichkeit. „Adäquanz“ kommt in mehrfacher Hinsicht Bedeutung zu, namentlich für die verwendete Technik und die Organisation.

[37] Eng verwoben mit „Adäquanz“ und von herausragender Bedeutung ist das Kriterium der Aktualität. Dieses Alleinstellungsmerkmal gegenüber herkömmlichen Zertifizierungssystemen ist durch eine stetige Anpassung der Prüfanforderungen an die rechtlichen, tatsächlichen und technischen Gegebenheiten, eine Aktualisierung der verwendeten Systeme und Prozesse und der Überprüfung des Sachverstands und der Unabhängigkeit der eingesetzten Prüfungspersonen sicherzustellen.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Workshop „Cyberwarfare“

Workshop „Cyberwarfare“

Hannes Federrath[1], Götz Neuneck[2]

Erstens sind inzwischen mehrere konkrete Fälle von staatlich organisierten Angriffen auf andere Staaten bekannt, zweitens haben spätestens die Snowden-Enthüllungen klar gemacht, dass neben der passiven und aktiven Informationsgewinnung insbesondere auch aktive Angriffe zur Beeinträchtigung von Informations- und Kommunikationssystemen geplant und durchgeführt wurden und werden.

Der Workshop „Cyberwarfare“ soll etwa Beiträge zu folgenden Fragen liefern: Inwieweit unterscheiden sich staatlich organisierte Cyberangriffe von denen krimineller Organisationen und/oder einzelnen Hackern? Welche konkreten Gefahren gehen von zwischenstaatlichen Angriffen aus? Was sind die technischen Mittel für derartige Angriffe? Welche Kosten und welche Schäden werden durch die Angriffe verursacht? Können Angriffe gegen nicht-demokratische Staaten ethisch gerechtfertigt sein? Welche außenpolitische Dimension geht von den Angriffen aus? Inwieweit tragen Cyberangriffe militärische Züge (Cyberwar)? Gibt es für diese Form der Cyberwaffen ähnliche Regulierungsmechanismen wie beispielsweise für chemische, biologische oder nukleare Waffen?

Alle eingereichten Workshopbeiträge wurden von mindestens drei Programmkomiteemitgliedern begutachtet.

Programmkomitee

Jürgen Altmann, Universität Dortmund
Matthias Englert, TU Darmstadt
Hannes Federrath, Universität Hamburg
Sandro Gaycken, European School of Management and Technology Berlin
Gerald Kirchner, Zentrum für Naturwissenschaft und Friedensforschung Hamburg
Hans-Jörg Kreowski, Universität Bremen
Ulrike Lechner, Universität der Bundeswh München
Götz Neuneck, Institut für Friedensforschung und Sicherheitspolitik Hamburg
Stefan Oeter, Universität Hamburg

[1] Universität Hamburg, Fachbereich Informatik, Arbeitsbereich Sicherheit in Verteilten Systemen, Vogt-Kölln-Str. 20, 22527 Hamburg, federrath@informatik.uni-hamburg.de
[2] Institut für Friedensforschung und Sicherheitspolitik an der Universität Hamburg, Beim Schlump 83, 20144 Hamburg, neuneck@ifsh.de

Deterrence theory in the cyber-century.

Lessons from a state-of-the-art literature review

Annegret Bendiek[1] und Tobias Metzger[2]

Abstract: Cyberattacks from a variety of perpetrators are constantly rising. To achieve restraint from attacks, deterrence theory has long been considered a valuable concept. How do criticisms of classical deterrence apply in this relatively new domain, and how does cyberdeterrence differ? Can offensive cyber capabilities be effective in deterring adversaries or must kinetic retaliation be "on the table"? Freedman differentiates deterrence-by-retaliation and deterrence-by-denial; immediate versus general; narrow versus broad; and central versus extended deterrence. The authors argue for denial and retaliation as complimentary parts of an overall strategy consisting of resistance, resilience and responses. Unlike nuclear deterrence, cyberdeterrence is not a game of great powers or that of nation-states alone and grey zones serve only rogue actors in the medium-term. Only the establishment of clear rules, similar to the Budapest Cybercrime Convention, enables effective responses to prevailing threats. Agreement on international rules on civilian critical infrastructures would move the superficial discussion on whether or not cyberattacks are legitimate to the more relevant debate on which targets are acceptable, providing clarity for the development of effective military strategy.

Keywords: Cyberdeterrence; Cyberwarfare; Cyberdefence; Deterrence theory; Cybersecurity; Freedman.

1 Introduction

Cyberwar is regularly invoked in journalistic, academic and even political discourse. Yet, apart from the U.S.-Israeli Stuxnet attack on Iran's nuclear facilities in 2010, no cyberattack has ever caused large-scale physical damage. UK Labour last year urged their government to consider the "growing threat of cyberwarfare"[3] and former Defence Secretary Leon Panetta repeatedly warned of the lurking threat of a "cyber Pearl Harbour".[4] While the Stuxnet attack remains the only instance of severe physical damage inflicted via cyberspace, the Internet's increasing pervasiveness and national development of military units bear the risk of militarization. Connecting growing parts of the industry, energy production and society to the Internet fosters economic growth, increases efficiency and, benefitting from the Internet's anonymity, furthers human rights. Nevertheless, U.S. think-tanker Jason Healey cautions that, as the Internet of

[1] German Institute of International and Security Affairs (SWP), Senior Associate in the Research Division EU/Europe, Ludwigkirchplatz 3-4, 10719 Berlin, annegret.bendiek@swp-berlin.org
[2] German Institute of International and Security Affairs (SWP), Research Assistant in the Research Division EU/Europe, Ludwigkirchplatz 3-4, 10719 Berlin, tobias.metzger@swp-berlin.org
[3] [MA14]
[4] [Se12]

Things (IoT) spreads, "a cyberattack will destroy not only ones and zeros, but things made of steel and concrete. And when they break, people will die."[5]

To achieve restraint from attacks, deterrence theory has long been considered a valuable concept. While deterrence will remain an instrument in security policy – any consideration of deterrence theory must acknowledge that its limitation to the military, and more specifically the nuclear, domain is insufficient. In line with this, various authors have discussed the extent to which deterrence theory is applicable to cyberspace. Their findings of appropriateness and limitations are a necessary starting point for policy-making recommendations. How do the criticisms of classical deterrence apply in this relatively new domain, and how does cyberdeterrence differ from its kinetic counterpart? Can offensive cyber capabilities be effective in deterring adversaries? Must kinetic retaliation be "on the table" for deterrence to succeed? Which changes are required for its implementation and where do key challenges lie? Unlike Gaycken and Martinelli, our definition of cyberdeterrence is built on both deterrence of cyberattacks and deterrence by threatening cyberattacks, arguing that rather than being separate, they are different escalatory steps and conceptually cannot be separated.[6] The means of deterrence are part of an overall toolbox and discussing cyber separately would be similar to speaking, for instance, solely about deterrence effects of the navy or air force. The authors build upon Lawrence Freedman's types of deterrence: deterrence-by-retaliation and deterrence–by-denial; "narrow" vs. "broad"; "central" vs. "extended" and "immediate" vs. "general" deterrence. Addressing these questions is essential for determining a deterrence strategy's effectiveness. In its study for the U.S. Air Force, RAND subsumes only questions of punishment as deterrence, while referring to all deterrence-by-denial mechanisms as "defence".[7] It suggests that deterrence-by-denial and deterrence-by-retaliation ought to be considered separately. We argue that they are complimentary. For deterrence to be effective in cyberspace, actors dealing with foreign and security policy can find useful reference points in the state-of-the-art literature on deterrence theory and cyberdeterrence.

2 In theory – Deterrence theory and cyberspace

Throughout the Cold War, deterrence theory was the preferred framework of analysis and of military doctrine to explain the influence of nuclear weapons, and to argue that nuclear powers, fearing the consequences, would not go to war with each other. Some authors have since applied the theoretical framework to cyberspace, as cyberdeterrence[8].

[5] Jason Healey in [SD14a], 31
[6] Cf. [GM13]: Differentiation made between „cybered deterrence", as deterrence by cyber means, and "cyber deterrence", as deterrence of cyberattacks.
[7] Cf. [Li09], 7,8
[8] For in-depth accounts, see [Kn10]; [Ci12]; and [GM13]

While both domains share characteristics, such as the offensive advantage, given the difficulty and costliness of defence[9], significant differences exist.

The emergence of deterrence in military theory dates back to the 1920s/30s when the first flight bombers were considered unstoppable by defensive measures. Then, strategists thought that large-scale attacks on one's cities could only be prevented, if the other side feared counter-attacks of similar or greater magnitude. The first nuclear bombs demonstrated a similar offensive advantage, and Bernard Brodie was among the first to observe that "from now on [the military establishment's] chief purpose must be to avert [wars]".[10] Deterrence theory gained prominence and developed to its present state during the Cold War nuclear stand-off between the USA and the Soviet Union. The term goes back to the Latin "dēterrere", meaning to "frighten from or away", and is defined as "to discourage and turn aside or restrain by fear".[11] "Deterrence is concerned with discouraging others from acting in ways that advantage them but harm you". This definition highlights the two notions of deterrence, firstly the would-be-attacker's turning back due to the other's defences, and secondly restraint for fear of retaliation. The threat of force by *'A'* and the voluntary restraint of *'B'* are central elements. As Freedman puts it: "strategies geared to coercing others to act in ways they might consider harmful but advantage you have been described as compellence or coercive diplomacy".[12] The main difference is that deterrence seeks to preserve the "status quo" and is limited to persuading someone not to do something, while compellence seeks to enforce a change, typically within a frame of urgency.

Deterrence requires two components: the expressed intention of *A* to defend an interest; as well as the ability to achieve the defence or the (perceived) certainty by *B* that interference with *A's* interest will be costly for the attacker, i.e. credibility. However, signalling – used to communicate both the interest to be defended and the threats to be implemented in case of non-compliance – is never straight forward. Freedman stresses the possibility of *A* badly articulating or *B* misunderstanding the threat, thus rendering deterrence ineffective. Signalling or "brandishing"[13] of weapon capability can be very costly, as evident in Israel's military operations aimed at deterring future attacks. The ultimate aim is strategic deterrence, thus creating "internalised deterrence", no longer requiring explicit signalling.[14] Robert Jervis describes an evolution of "three waves of deterrence": Firstly, Brodie's concept of deterrence to avert wars when only the West possessed deployable nuclear weapons; secondly, the rise of the Soviet Union as a nuclear power leading to a bipolar world evoking the question "if nuclear war couldn't be fought how could it be threatened?". At this stage second-strike capabilities deployable from submarines assured Mutually Assured Destruction (MAD), making it

[9] See also [Li14], 5, 96: Calculations put offense at 132 times cheaper than defence, and estimate a $100m "cyber army" as being able to overcome any U.S. cyberdefence, and cf. [Ma12]
[10] [Br46], 31 and cf. [Fr04], 9-11
[11] Search for "to deter", [Ox14]
[12] Both quotes [Fr04], 109
[13] Cf. [Li13]
[14] Cf. [Fr04], 28-32

impossible to inflict sufficient damage to disarm an enemy, which would prevent retaliatory attacks. Game theory was used to evaluate whether cost-benefit evaluations could still favour deterrence postures, issuing a "threat that leaves something to chance"[15] or threatening sanctions using limited (non-nuclear) strikes. U.S. State Secretary Dulles argued for deterrence of would-be-aggressors from a cost-efficiency perspective, calling it "the way to getting maximum protection at bearable cost".[16]

Jeremy Bentham, one of the inventors of deterrence theory in criminology assumes rational individuals capable of performing cost-benefit calculations prior to taking action. The third wave raised considerable doubts about this rational actor model, an important pillar of deterrence theory, arguing that groupthink, misperceptions and bureaucratic politics often overruled mere cost-benefit calculations. Rationality is subjective, and the challenge of deterrence signalling lies in identifying the opposing side's rationale.[17] Furthermore, cost-benefit calculations require clarity and predictability of sentencing and proportionality between punishment and violation. Bentham observed that where arrest is unlikely, severe punishment can help keep up a criminal's expected cost of getting caught, but once punishment is perceived as disproportionate, it loses its effect.[18] In line with the controversy of "rational" actions, signalling can be misunderstood given differences in culture or education. Furthermore, there are claims of fallacies in traditional deterrence theory's sole consideration of state actors. Third-wave theorists Alexander George and Richard Smoke disapproved of the exaggerated role of the military vis-à-vis other foreign policy tools, especially positive inducements.[19] The rise of non-state actors, "rogue states" and "terrorists", brought a possible fourth wave. These actors may be beyond containment by deterrence since they have no nation or citizens to defend. Finally, theorists introduced the differentiation between interest-based and norms-based deterrence in the 1990s to overcome previous challenges. The former aims at deterring challengers of "hard" national interests, while the latter advocates less clear-cut norms. Freedman argues for norms-based deterrence to reinforce "certain values to the point where it is well understood that they must not be violated".[20]

Deterrence-by-retaliation and deterrence-by-denial

To dissuade would-be-offenders from attacking, deterrence theory typically distinguishes two means: denial and retaliation. In the original criminological context, Bentham describes deterrence-by-retaliation, or deterrence-by-punishment, as imposing the "significant likelihood of any culprit being apprehended, brought to trial, found guilty and then receiving a sentence … that will make an impression not only on their future behaviour but on the behaviour of others".[21] This was the core idea of nuclear deterrence until the 1970s when U.S. President Reagan's "Strategic Defense Initiative"

[15] [Sc66] and cf. [Po08]
[16] [Du89] as quoted in [Fr04], 9
[17] Cf. [Mo77]. Bentham's collected works are available online, cf. [Be43]
[18] Cf. [Li09], 29
[19] Cf. [GS74]
[20] [Fr04], 4
[21] [Fr04], 60, 61

(SDI) introduced the notion that is was better "to protect than avenge".[22] The SDI suggested a deterrence-by-denial approach. Deterrence-by-resistance and deterrence-by-resilience are two different approaches within this concept. Resilience is the ability to quickly restore the original shape after an attack, Quick recovery limits potential gains and can convince an opponent not to attack, if the cost of attacking becomes excessive. Both – resistance and resilience – are aimed at denying a would-be-offender's gains, either by building unsurmountable defence structures or by ensuring quick recovery following an attack. Freedman extends the distinction beyond, firstly, the denial versus retaliation debate: He, secondly, introduces "narrow" versus "broad" deterrence, distinguishing whether a particular military operation (e.g. the use of nuclear missiles) or any type of attack is to be deterred; thirdly, "central" versus "extended", thus including the protection of third-party allies in one's deterrence demeanour; fourthly, "immediate" versus "general" deterrence, distinguishing between a crisis situation between known actors and non-state of emergency deterrence against unknown would-be-aggressors.[23]

Many of the same elements apply regarding attacks through cyber means although with some limitations. Signalling, for one, faces comparable challenges, although further complicated by the multitude of actors including non-state groups and individuals. Other challenges are entirely new: In deterrence-by-retaliation, credibility is difficult to establish since demonstrating cyberpower and retaliating immediately and repeatedly is problematic, as outlined below. Although the U.S. and NATO have emphasized their willingness to respond to cyberattacks at a time, place and by means (including kinetic) of their choosing, there can hardly be automaticity in response. The similarity of nation-state and criminal cyberattacks requires time-consuming and costly forensics and close coordination between law enforcement and the military. With their reluctance of admitting to their development of offensive capabilities, and of strategically discussing their legality, Germany and other nations create opaqueness and significant grey areas. Given the offensive advantage, the number of attackers using cheap, readily available tools will continuously rise, empowering non-established powers such as Iran, North Korea or even Daesh/IS. Reversing this trend requires getting serious about agreeing on international norms and improving both defences, especially employees' and citizens' "cyber-hygiene", and about enforcement. Lack of clarity and impunity for attackers is a major roadblock for effective deterrence strategies.

3 In practice – Suitability of cyber: lessons and implications

The U.S. Army analysis that "[f]undamentally, there is no difference between deterrence in the cyber domain than in any other domain"[24] is flawed in at least three regards: First, cross-border differences in law enforcement and legal practice as well as unwillingness to cooperate allow attackers to act with impunity, diminishing the deterrent's credibility.

[22] [Fr04], 19
[23] Cf. [Fr04], 32-42
[24] [Ph13], 16

Second, since "cyberweapons" rely largely on previously unknown, so called zero-day, vulnerabilities and cannot be displayed prior to their use, it is difficult to demonstrate power. Third, deterrence-by-denial differs greatly, since in cyber "you have to work from the assumption that your networks are already compromised"[25], meaning deterrence is constantly failing. In the nuclear context, all intrusions must be deterred, and a single instance of failed deterrence could mean the use of nuclear warheads and large-scale loss of life. Fourth, different from nuclear confrontation, uncertainty arises from the multitude of actors which threaten harm to one's systems and from the difficulty of quickly attributing an attack. Fearing unforeseen escalation, this may hinder immediate retaliation, especially when retaliating with kinetic strikes.

Ambiguously defined interests, misperceived signalling and uncertainty on how to demonstrate force and how to respond, hamper deterrence postures. Melissa Hathaway, former director of a classified high-level effort to establish a U.S. cyberdeterrence strategy, admitted that "we didn't even come close".[26] While being a useful frame of analysis, cyberdeterrence fails to satisfy any of Patrick Morgan's six elements of classical deterrence theory, according to Stevens.[27] Libicki aptly summarizes the core issues to be observed in national cybersecurity efforts: "The ambiguities of cyberdeterrence contrast starkly with the clarities of nuclear deterrence. In the Cold War nuclear realm, attribution of attack was not a problem; the prospect of battle damage was clear; the 1,000th bomb could be as powerful as the first; counterforce was possible; there were no third parties to worry about; private firms were not expected to defend themselves; any hostile nuclear use crossed an acknowledged threshold; no higher levels of war existed; and both sides always had a lot to lose."[28]

3.1 Key challenges: Credibility and capability to display and use force

Beyond technical issues – which are being addressed – deterrence-by-retaliation is a question of credibility and capability. Firstly, could governments effectively deploy cyberforce to respond to attacks, and should they use kinetic force – and if so under which circumstances – to react to cyberattacks, potentially risking escalation? Secondly, should there be stronger declarations of cyberforce to signal one's ability? For deterrence-by-retaliation to work, the capacity to display force is crucial. Only verifiable tests and the demonstration of nuclear weaponry's destructive force convinced nations that future use ought to be feared. Similar demonstration of cyberpower[29] is unlikely because of the "one use only component", meaning that any use reveals significant information necessary to defend against future attacks. Deterrence-by-retaliation requires

[25] Michael Daniel, Cybersecurity Coordinator of U.S. President Obama, at [SD14b]
[26] Hathaway, Melissa quoted in [Ma10]. On cyberdeterrence in U.S. strategy, see [St12], 154
[27] Cf. [Mo03], 8 and cf. [St12], 152: 1. No prevailing military conflict; 2. rational choice models differ for non-state actors; 3. the challenge of attribution complicates immediate retaliation; 4. repeated retaliation and certainty of inflicting severe pain is hampered; 5. the difficulty of demonstrating offensive capabilities lessens credibility; and 6. a multitude of (non-state) actors constantly threatens stability, risking escalation.
[28] [Li09], xvi
[29] For cyberpower's relevance for policy-making, see [Sh11]

a demonstration of force, but most cyberweapons are rendered useless if vulnerabilities are closed.[30] Signalling is complicated by the multitude of actors beyond the great powers, and including the private sector and individuals. Willingness to retaliate requires unmistakable signalling on the interest to be defended. While offensive capabilities have been revealed, Germany has never publicly proclaimed their existence or threatened adversaries with their use in response to attacks. France and the UK publicly announced offensive capabilities, and France stated its intent of becoming a cyberpower in its 2011 cyber doctrine, but both left it unknown when and how cyberweapons could be used.[31] The U.S., on the other hand, emphasized that their response to a cyberattack would not necessarily be via cyberspace, but that all options would be on the table.[32]

Mutually assured destruction, equally, is out of the question, changing the cost-benefit calculation in favour of attack. If the MAD principle does not apply, the consequences of retaliation are considerably less severe, reducing the inherent costs of an attack. This leads to an advantage for the party striking first, according to offence-defence theory, making attacks more likely. While this suggests for other characteristics of offence-defence theory to apply – e.g. that an arms race ensues – the lacking ability to make quick and decisive victory changes the equation. If cyberattacks cannot disarm an opponent, what is the point of rushing into retaliation? An attack may be stopped, but the attacker cannot be disarmed since cyberattacks can be conducted from third-party hardware, including internet cafés, unsecured wireless networks or infected computers, as part of a botnet. Furthermore, the asymmetric nature of cyberspace is a common argument against retaliatory attacks against criminals or "cyber-terrorists". Deterrence fails if there is no valid target to strike back at. The less connected an adversary, the less vulnerable he is to retaliation with cyber means. Retaliation by kinetic means, on the other hand, bears the risk of incurring injuries or deaths where the initial attack did not. Retaliation "requires not only breaking into sufficiently privileged levels, but also figuring out how to induce a system to fail and keep on failing".[33] The ability to induce superficial damage, which is quickly repaired, has no deterrent effect. However, predicting damage ahead of an attack is difficult, as this depends on the other's technical and procedural resilience but also on chance – e.g. whether patches are installed, closing previously existing backdoors. The half-life of exploits creates a "use-it-or-lose-it dilemma".[34] While initial attacks allow for intensive intelligence work, repeated retaliation is costly, if not impossible, since it requires firstly rapid attribution and secondly immediate and continuous knowledge of the target system. Figure two puts this question in context.

[30] Cf. [Li13], xv-xvii and vii-viii
[31] Cf. [Bl13], and [Li13], 25
[32] Cf. [De15]
[33] [Li13], viii
[34] [Li09], 58

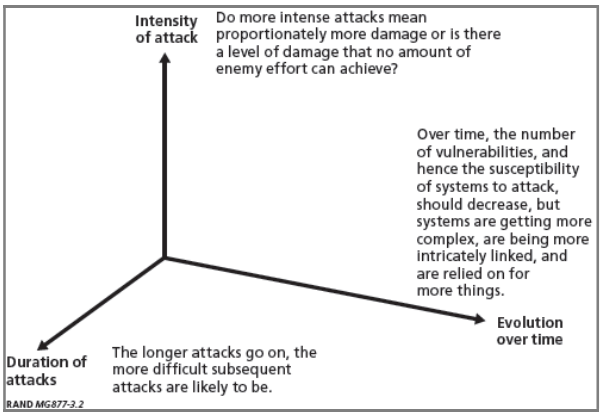


Abb. 1: Limits to retaliation in cyberspace[35]

Additionally, once malware source code such as that of Stuxnet became available online, it revealed flaws in Siemens' operating system, thus enabling their use against other CI. Attackers "may risk handing over ammunition to the enemy as a blueprint for the latter to develop a cyber weapon of its own", as a result of reverse engineering.[36] Instead of taking advantage of such flaws and even buying such information off the black-market, governments must inform the manufacturer to rid systems of backdoors. NSA exploitation of zero-day vulnerabilities and their purchasing with a 2015 budget of $25 million dollars[37] strengthens this business, and may open "Pandora's Box" encouraging others to follow suit.

3.2 How to deter? Deterrence-by-denial and deterrence-by-retaliation

In a cyberspace of easy targets, weak cross-border cooperation and diverging legislation the odds are tilted in favour of attackers. As RAND puts it, "if cyberattacks can be conducted with impunity, the attacker has little reason to stop."[38] Former White House advisor Richard Clarke admits that currently "we cannot deter other nations with our cyber weapons. Nor are we likely to be deterred". According to him, deterrence requires getting "serious about deploying effective cyber defences for some key networks" and "since we have not even started to do that, deterrence theory … plays no significant role in stopping cyber war today". He demands a defensive triad of improving backbone network security, protecting critical infrastructures (CI) and improving the security of military networks and weapons.[39]

Determining the type of defence

[35] [Li09], 60
[36] [Sh14], 78
[37] Cf. [Gr14]
[38] [Li09], xvi
[39] Cf. [CK12]

For deterrence-by-denial to succeed, cost-benefit calculations of attackers must turn negative: "A strong defense deters an attack by convincing an attacker there will be no gains commensurate with the cost of attack".[40] Good defences make an attack less likely to succeed, add credibility to retaliatory measures, reduce the success rate of low-sophistication third-party attacks and make it easier to attribute attacks. As an important differentiation, we note that deterrence-by-denial exists in the pre-event as defence, and in the post-event as resilience. Bologna et al. urge to move from a "fortress" to a "resilience" approach[41], while a RAND study suggests "that, in this medium, the best defense is not necessarily a good offense; it is usually a good defense".[42] The importance of defence is evidenced in NATO's emphasis on cyber-hygiene. If a system is ridded of low-level disturbances, advanced threats are easier to identify and to counter.[43] "Honeypots", entirely sterile systems which capture malware and allow analysts to observe their behaviour in a vacuum, leverage this observation. Another feature of cybersecurity strategies – especially that of the EU – is "strategic dependency management".[44] This includes all measures to secure key components of the supply chain by deciding which levels of strategic independence are required in industry R&D, manufacturing and maintenance of important IT components. It serves to minimize national risks, but often faces accusations of government protectionism. After the Snowden revelations risk management strategies led China to reduce their reliance on U.S. technology, e.g. deciding against Microsoft's Windows 8.[45] It also led to calls within Europe for greater technological sovereignty from U.S. communication infrastructure. Such measures, while sensible in light of deterrence-by-denial, reverse the purpose of the global Internet and defensive measures, e.g. Iran's creation of parallel structures for a national network to prevent high-level cyberattacks, can be bypassed as evidenced in "Stuxnet", which infected the nuclear reactor's software via a USB-stick. While Germany uses offensive means to stop ongoing attacks[46] and works to strengthen cross-border criminal law enforcement, most actions can be seen from a deterrence-by-denial angle. The strong defence network of public and private Computer Emergency Response Teams (CERT) is one component thereof. However, distinguishing between offence and defence is not always easy since evaluating one's own defences requires "penetration-testing" and thus the ability to intrude into systems.

Adding offence to the equation

Other experts strongly disagree with the emphasis on defence, arguing that "in an offense-dominant environment, a fortress mentality will not work".[47] In the nuclear context, deterrence-by-denial approaches were largely inapt given the unbearable costs

[40] [Ph13], 2-4
[41] Cf. [Bo13]
[42] [Li09], 176
[43] Interview [NA13]
[44] Interview [ED14]
[45] Cf. [Re14]
[46] Interview [BM14]
[47] [Ly10]

of defensive failure. In cyberspace, lack of clarity and credibility of punishment encourages cyberattackers to test defences and push their limits, defying deterrence-by-denial. The low cost, as the sum of conducting the attack plus the cost of likely penalties, furthers the offensive advantage, as does the existence of a vast black market, offering anything from zero-day exploits to off-the-shelf services to conduct denial-of-service attacks.[48] The nature of software development renders entirely avoiding bugs, or loopholes impossible, and quickly responding to attacks and closing vulnerabilities entails significant costs for the defender.

3.3 When and whom to deter? Immediate vs. general deterrence and the challenge of attribution

Immediate vs. general deterrence is the question of whether deterrence is aimed at a specific adversary during an ongoing conflict or more generally at any would-be-offender during peacetime. Due to its development time and the need to know a system's vulnerabilities, cyberforce is little appropriate as a concrete deterrence measure once a conflict has erupted. On the other hand, retaliation against cyberattacks cannot be adhoc because of the need for time-consuming forensics. Thus, both general deterrence strategies aiming to deter any attack by cyberforce as immediate deterrence by cyber means alone are improbable. Rather cyber can be an instrument in the broader toolbox. "Active defence" is ill-fitted for various technical, political and legal reasons. Such fully-automated retaliatory attacks may damage computers that are part of botnets without their owner's knowledge, while the botnet as a whole will simply replace the neutralised computer. Furthermore, nation states are not likely to legalise such mechanisms, thereby allowing private sector "corporate vigilantisms" to violate the government's monopoly on the use of force. The multitude of actors entails further difficulties: During conflicts, patriotic and other hackers can engage in disruptive activities to add an unpredictable step of escalation. Governments may not be able to control these groups, weakening the implied promise of deterrence that "if you stop, we stop".[49] Thus, cyber threats are only credible in combination with traditional, threats. Figure 4 introduces one possible model adding levels of retaliation via cyberspace: Firstly, escalation beyond political and economic sanctions using low-level cyberattacks and, secondly, escalating kinetic strikes ahead of the use of tactical nuclear warheads.

[48] Cf. [Sh11], 97 and Interview [ED14]
[49] Cf. [ED14] and [Li], 62

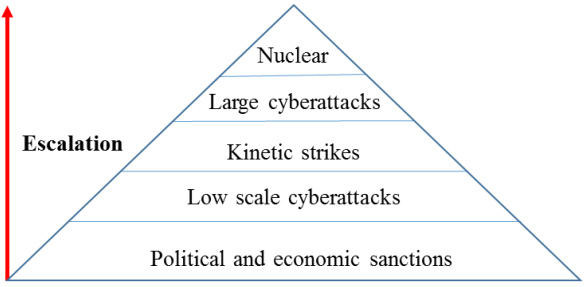


Abb. 2: A possible model of escalation[50]

Since knowing the opponent is crucial in designing defensive or offensive responses, attribution is core. At the moment of an incident, it is difficult to assess whether one is dealing with a technical accident, indiscriminately spreading malware, or a targeted attack. For deterrence-by-retaliation to work, aggressors must be convinced that they will be identified and punished; and attribution must be bulletproof to avoid creating new enemies and to convince third parties that the retaliatory measure is not an attack in itself. A 2013 NATO CCDCOE study lists anonymity as "[denying] identification of malicious actors, thus making deterrence policies futile, the undertaking of diplomatic, political and economic reaction measures difficult, and the application of legal remedies, e.g. countermeasures, impossible".[51] Still, "the attribution problem is partly artificial: As analyses of individual indictments following cyberattacks – both on the U.S. in 2013 and 2014 and Estonia in 2007 – demonstrate, attribution is costly and time-consuming, but possible. It combines technical measures and intelligence operations to pinpoint the culprit. Dealing with state actors, malware source code and the programming style can be mapped against previous incidents. Such forensics revealed similarities between Stuxnet and Flame, intensifying accusations of U.S.-Israeli involvement in both high-complexity tools.[52] Both long-term rivals of Iran were singled out long before Snowden provided proof. The major difficulty lies in prosecution once perpetrators are identified. In the Estonian case, Russia refused to prosecute indicted citizens, saying that their DDoS attacks did not amount to legal violations in Russia.[53]

3.4 What to deter? Narrow vs. broad deterrence

In narrow vs. broad deterrence the core question is "What is to be deterred?". In the current context, and based on initial definitions of operational and strategic cyberwar, a war of cyberattacks is highly unlikely. "Narrow" deterrence-by-retaliation requires unmistakable positioning as to what is acceptable and what is not. The outcomes of U.S. political cyberespionage and simultaneous indictment of Chinese officers for economic

[50] Own figure based on Interview [ED14]
[51] [Zi13], XVI
[52] Interview [NA14] and Interview [EE14]
[53] Interview [NA14]

cyberespionage may set an important precedent. The U.S. draws somewhat artificial lines between political and economic cyberespionage, indicting Chinese officers for industrial espionage while itself facing accusations of spying on OPEC and European arms manufacturers, hacking China's Huawei and supporting, or at least tolerating, British hacking into Belgium telecommunications giant Belgacom.[54] The use of cyberattacks such as Stuxnet risks escalation, given the lack of international agreement on what is acceptable. U.S. President Obama issued an executive order on 1 April 2015 calling for sanctions against individual and organised cyberattackers, especially those conducting attacks on critical infrastructure.[55] This is an important starting point.

NATO CCDCOE's Tallinn Manual[56] insisted on the applicability of Geneva Convention IV on the protection of civilians in cyberspace. This prohibits disproportional attacks against civilian and non-military installations. However, some discard specific "Cyber Geneva Conventions" as unverifiable and unlikely to yield results in the short- to medium-term, insisting instead on internationally agreed principles.[57] The UN's ITU sees hope in "effective norms against cyberaggression […] reining in unacceptable forms of behaviour".[58] International norms on restraint in cyberwarfare appear the only hope to resolving diplomatic upsets about cyberespionage and cyberattacks. The currently prevailing grey zones ultimately only serve non-state and criminal actors. As a first step, nations worldwide must determine what constitutes an illegal attack and warrants responses from the international community. The protection of civilian critical infrastructure should be among the leading goals of international efforts. The UN General Assembly and its working groups on cyber are the most appropriate forum to discuss which targets and which weapons are to be prohibited. Military and political strategists can thereafter develop a narrow deterrence regime against specific actors, to protect certain systems and to single out weapons to be deterred. Cyberpower cannot deter any aggression, but deterrence can be rendered effective by outlining retaliatory measures in case of cyberattacks against a narrow set of targets or using a narrow set of attack mechanisms.

3.5 For whom? Central vs. extended deterrence

The question of central vs. extended deterrence extends beyond whether EU and NATO allies possess collective defence mechanisms, in their respective Article 222 of the Treaty on the Functioning of the European Union and Article 5 of the North Atlantic Treaty. For extended deterrence to be effective, allies would have to signal to would-be-offenders that violations against one ally will result in a joint response. While the articles require considerable thresholds – the loss of human life arguably being one – the increasing connectedness of e.g. energy grids increases the threat for neighbours during

[54] Cf. [Fe14], cf. [Sp13a] and [Sp13b]
[55] [Th15]
[56] See [Sc13]
[57] Interview [EE14]
[58] Cf. [Ha10]

cyberattacks against one country. NATO doctrine recognizes cyber as part of NATO's collective defence and emphasises that retaliation may be "by any means necessary".[59] Similar to any public-private cooperation, however, the contentious item is granting access to one's networks. Public-private and civilian-military cooperation as well as inter-state collective cyberdefence requires a great deal of trust, since technical assistance can only be provided after having been granted wide-ranging access, possibly revealing additional vulnerabilities.[60] Governments may, furthermore, consider it little desirable to perform operational tasks in securing privately-owned critical infrastructure, which may lead to free-riding and insufficient investment in cybersecurity. Given this unwillingness to grant access, NATO-coordinated joint cyberattacks are unlikely. All nations have to, first and foremost, secure themselves. Technologically leading NATO allies should engage in more strategic dialogue with their partners to increase overall preparedness. This can then be the basis for extended deterrence. President Obama strengthened the idea of such concepts, stating that "the United States will respond to hostile acts in cyberspace as we would to any other threat to our country. … We reserve the right to use all necessary means – diplomatic, informational, military, and economic – as appropriate and consistent with applicable international law, in order to defend our Nation, **our allies, our partners**, and our interests."[61]

4 Conclusion and outlook

Previous academic research on deterrence theory and its application to cyberspace, leads to a series of conclusions for policy-making. Theorist and UK foreign policy advisor Lawrence Freedman differentiates 1) deterrence-by-retaliation and deterrence-by-denial; 2) immediate vs. general; 3) narrow vs. broad; and 4) central vs. extended deterrence. Taking into consideration the extensive literature review, cyberdeterrence strategies ought to be inclusive of both retaliatory and denying mechanisms. This chiefly includes four elements: Firstly, resistance by developing strong guidelines against voluntary and accidental disruptions; secondly, resilience to quickly and fully recover from attacks; thirdly, the definition of clear-cut global rules on acceptable practice in peacetime and on the legality of targets during armed conflict; finally, a national strategy of responses within these rules, ranging from criminal prosecution and political condemnation to (economic) sanctions and, finally, measures of active defence and retaliation. Any retaliatory strategy must also address the question of whether kinetic strikes are warranted in the case of sincere damage to the property and lives of citizens.

As "cyberweapons" require target-specific development, quick retaliatory cyberstrikes are impossible. Therefore, and due to the fact that attribution requires time-consuming forensics, retaliation with cyber means cannot be done ad-hoc. Defence and offence must thus be joined into a broader deterrence strategy. Those responsible for national defence

[59] Interview [NA14]
[60] Cf. [Li09], vxiii and Interview [NA14]
[61] [Th11], 14

should focus on protecting their own networks, most importantly communication and weapons control. This will add credibility to any future deterrence strategy. However, deterrence strategies should start not merely from the defence community, but hand-in-hand with activities of the broader government. Currently, rather than strategically thinking about how to tackle threats, authorities are still trying to understand threat agents and vulnerability landscapes. Classical deterrence requires defining interests and drawing redlines. Governments must clarify their interests and prioritize what is to be protected before deterrence postures can be effective. Which services and infrastructures are critical for the country's functioning? Which actors can ensure their security? Purely focusing on the military is inappropriate, and while the military is good at defending its own networks, it has hardly any experience in cooperating with the public and, even less so, the private sector. This, however, is an essential part of cybersecurity. To co-opt the private sector, governments have to create win-win-situations rather than imposing from above, and recognize that, unlike nuclear deterrence, cyberdeterrence is not a game of great powers and not even that of nation-states alone. Neither the European approach of imposed cooperation, as in the draft NIS Directive's mandatory reporting, nor the U.S. approach of largely leaving cybersecurity to market forces appear sufficient. Germany's National Cybersecurity Council (NCSR) and National Cyberdefence Centre (NCAZ) may provide starting points of such integration.

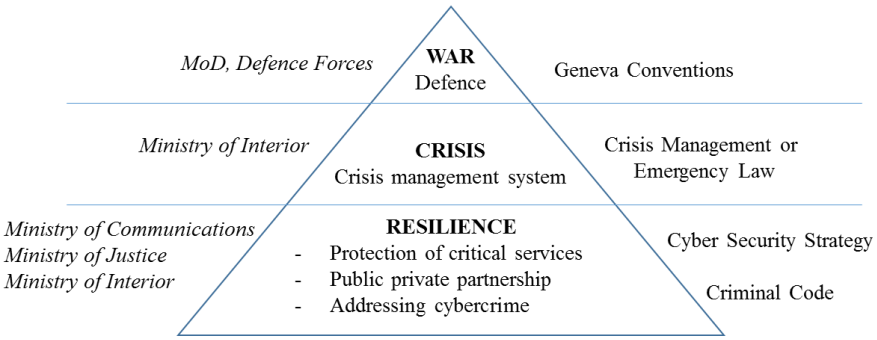


Abb. 3: EEAS figure on a possible inter-ministry division of labour[62]

Only the establishment of clear rules on acceptable practice in cyberspace will enable broad and effective responses to prevailing threats. Nations and individuals offending such treaties would then feel the deterring threat of retaliation with sanctions and criminals could be targeted with legal prosecution across national borders. The current grey zones in cyber conduct serve no one in the medium term. International agreements, similar to the Budapest Convention on Cybercrime[63], are crucial to preserving the internet as an engine of the global economy. The introduction of norms-based deterrence may thus provide a useful framework. Since cross-border cooperation is crucial, nation-states have to agree on acceptable thresholds, e.g. regarding the legality of

[62] Own figure based on [Ti12], 5
[63] Cf. [Co01]

cyberespionage. There is a sense of urgency in statements of high-ranking military and political leaders, acknowledging the risks of making cars, hospitals, energy grids and even prisons controllable via the Internet. Reluctance to tackle these questions politically will complicate matters, as the private sector engages in politics and vigilantism by striking back against botnets or by independently acting against censorship. Governments would most effectively strengthen the private sector by establishing legal clarity.

In order to achieve agreement on norms, it is sensible to build upon existing regimes such as the Geneva Conventions, establishing that offences through cyberspace warrant similar consequences as their traditional counterparts. The NATO CCDCOE's Tallinn Manual and its Peacetime Regime for State Activities in Cyberspace are important and should be seen as the basis for future efforts. In a first step, definitions on civilian critical infrastructures should be agreed upon internationally and undertaking or supporting attacks against such must be prohibited. This would move the discussion on the legitimacy of cyberattacks to the more relevant debate on which targets are acceptable, providing clarity for the development of effective military strategy. To achieve much-needed global progress, international disagreements over the appropriate governance of the internet or the provision of freedoms online should be dealt with separately.

List of references

[Be14] Bendiek, Annegret. “Tests of Partnership: Transatlantic Cooperation in Cyber Security, Internet Governance, and Data Protection.” SWP - German Institute for International and Security Affairs. March 2014.

[Be43] Bentham, Jeremy. The Works of Jeremy Bentham, 11 volumes. Edited by John Bowring. Vols. Edinburgh: William Tait, 1838-1843.

[Bl13] Blitz, James. “UK becomes first state to admit to offensive cyber attack capability.” Financial Times. 29 September 2013.

[BM14] BMVg Cybersecurity Expert, interview by Tobias Metzger. (24 June 2014).

[Bo13] Bologna, Sandro, Alessandro Fasani, and Maurizio Martellini. “From Fortress to Resilience.” In Cyber Security: Deterrence and IT Protection for Critical Infrastructures, edited by Maurizio Martellini, 53-56. Heidelberg: Springer, 2013.

[Br46] Brodie, Bernard, ed. The Absolute Weapon. New York, NY: Harcourt Brace, 1946.

[Ci12] Cilluffo, Frank J., Sharon L. Cardash, and George C. Salmoiraghi. “A Blueprint for Cyber Deterrence: Building Stability through Strength.” Military and Strategic Affairs, Vol. 4, No. 3, 2012: 3-23.

[CK12] Clarke, Richard A., and Robert Knake. Cyber War: The Next Threat to National Security and What to Do About It. New York: Ecco, 2012.

[Co01] Council of Europe. Convention on Cybercrime - Treaty Doc. 108-11, ETS No. 185. 23 November 2001.

[De15] Department of Defense. The DoD Cyber Strategy. Document released on 17 April 2015. www.defense.gov/home/features/2015/0415_cyber-strategy

[Du89] Dulles, John Foster. “The Evolution of Foreign Policy” In US Nuclear Strategy: A Reader, edit. by P. Bobbit, L. Freedman and G. Treverton. London: Macmillan, 1989.

[ED14] EDA Cybersecurity Official, interview by Tobias Metzger. (24 June 2014).

[EE14] EEAS Cybersecurity Official, interview by Tobias Metzger. (27 June 2014).

[Fe14] Ferranti, Marc. “NSA hacked into servers at Huawei headquarters, reports say.” Computerworld. 23 March 2014.

[Fr04] Freedman, Lawrence. Deterrence. Cambridge: Polity Press, 2004.

[GM13] Gaycken, Sandro, and Maurizio Martellini. “Cyber as Deterrent.” In Deterrence and IT Protection for Critical Infrastructures, edited by Maurizio

Martinelli, 1-10. Heidelberg: Springer, 2013.

[Gr14] Grossman, Lev. “Inside the Code War.” TIME, 2014: Vol. 184, No. 3: 20-27.

[GS74] George, Alexander L., and Richard Smoke. Deterrence in American Foreign Policy: Theory and Practice. New York: Columbia University Press, 1974.

[Ha10] Harknett, Richard J., John P. Callaghan, and Rudi Kauffman. “Leaving Deterrence Behind: War-Fighting and National Cybersecurity.” Journal of Homeland Security & Emergency Management, 2010.

[Kn10] Knopf, Jeffrey W. “The Fourth Wave in Deterrence Research.” Contemporary Security Policy, Vol. 31, No. 1 April 2010: 1-33.

[Li09] Libicki, Martin C. “Cyberdeterrence and Cyberwar - Prepared for the United States Air Force.” RAND Corporation. 2009.

[Li13] –. “Brandishing Cyberattack Capabilities.” RAND National Defense Research Institute. 2013.

[Li14] Lieber, Keir. “The Offense-Defense Balance and Cyber Warfare.” In Cyber Analogies, edited by Emily O. Goldman and John Arquilla, 96-107. Monterey: Naval Postgraduate School and U.S. Cyber Command, 2014.

[Ly10] Lynn III, William J. “Defending a New Domain: The Pentagon's Cyberstrategy.” Foreign Affairs Journal. September/October 2010.

[Ma10] Markoff, John, David E. Sanger, and Thom Shanker. “In Digital Combat, U.S. Finds No Easy Deterrent.” The New York Times. 25 January 2010.

[Ma12] Malone, Patrick J. Offense-Defense Balance in Cyberspace: A Proposed Model. Monterey: Naval Postgraduate School, 2012.

[Ma14] Mason, Rowena. “Labour urges strategic defence review to consider cyberwar threat.” The Guardian. 24 March 2014.

[Mo03] Morgan, Patrick. Deterrence Now. Cambridge: Cambridge University Press, 2003.

[Mo77] –. Deterrence: A Conceptual Analysis. Beverly Hills, CA: Sage Publications, 1977.

[NA13] NATO Cybersecurity Official, interview by Tobias Metzger. (30 April 2013).

[NA14] NATO Cybersecurity Official, interview by Tobias Metzger. (17 June 2014).

[Ox14] Oxford English Dictionary. Oxford English Dictionary online - "deter". 2014.

[Ph13] Philbin, Lt. Col. Michael J. “Cyber Deterrence: An Old Concept in a New Domain.” U.S. Army War College: Strategy Research Project. March 2013.

[Po08] Powell, Robert. Nuclear Deterrence Theory: The Search for Credibility. Cambridge: Cambridge University Press, 2008.

[Re14] Reuters. China bans use of Microsoft's Windows 8 on government computers.

20 May 2014.

[Sc13] Schmitt, Michael N., ed. Tallinn Manual on the International Law Applicable to Cyber Warfare. New York: Cambridge University Press, 2013.

[Sc66] Schelling, Thomas. Arms and Influence. New Haven: Yale University Press, 1966.

[SD14a] SDA. “Annual conference.” Conference Report: Overhauling transatlantic security thinking. 15 July 2014.

[SD14b] —. Evening debate: Critical infrastructure protection in the cyber-age. Brussels. 30 June 2014.

[Se12] Secretary of Defence Panetta, Leon E. Remarks by Secretary Panetta on Cybersecurity to the Business Executives for National Security, New York City. 11 October 2012.

[Sh11] Sheldon, John B. “Deciphering Cyberpower: Strategic Purpose in Peace and War.” Strategic Studies Quarterly, Summer 2011: 95-112.

[Sh14] Shaheen, Salma. “Offense-Defense Balance in Cyber Warfare.” In Cyberspace and International Relations: Theory, Prospects and Challenges, edited by Jan-Frederik Kremer and Benedikt Müller, 77-94. Heidelberg: Springer, 2014.

[Sp13a] Spiegel. “Belgacom.” Belgacom Attack: Britain's GCHQ Hacked Belgian Telecoms Firm. 20 September 2013.

[Sp13b] —. “OPEC.” Oil Espionage: How the NSA and GCHQ Spied on OPEC. 11 November 2013.

[St12] Stevens, Tim. “A Cyberwar of Ideas? Deterrence and Norms in Cyberspace.” Contemporary Security Policy, Vol. 33, No. 1 April 2012: 148-170.

[Th11] The White House. International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World. May 2011.

[Th15] The White House. Executive Order. "Blocking the Property of Certain Persons Engaging in Significant Malicious Cyber-Enabled Activities". 01 April 2015.

[Ti12] Tiirma-Klaar, Heli. “Presentation: National Cyber Security Strategies and International Cyber Policy.” European External Action Service. 22 October 2012.

[Zi13] Ziolkowski, Katharina, ed. Peacetime Regime for State Activities in Cyberspace. International Law, International Relations and Diplomacy. Tallinn: NATO CCDCOE, 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Neue Ansätze für die Rüstungskontrolle bei Cyber-Konflikten

Ingo Ruhmann[1]

Abstract: Mit den Enthüllungen von Edward Snowden begann 2013 in Deutschland eine Debatte um die Überwachung der Telekommunikation. Die systematische Analyse der Dokumente zeigte jedoch von Beginn an, dass diese Überwachung nur ein Ausgangspunkt für die Kompromittierung der Sicherheit von IT-Systemen ist. Darin nehmen die USA zwar eine Sonderrolle ein, gut einhundert andere Staaten versuchen jedoch, es US-Akteuren bei der von Geheimdiensten, Militärs und klandestinen Gruppen in staatlichem Auftrag betriebenen Manipulation von IT-Systemen geleichzutun. Für eine auf ihre IT-Systeme angewiesene Informationsgesellschaft ist das im Vergleich zur Überwachung bedrohlichere Problem die Manipulation von IT-Systemen durch staatliche Akteure - Information Warfare. Eine Systematisierung der Erkenntnisse aus jüngerer Zeit über die dabei eingesetzten Werkzeuge erlaubt neue Ansätze für die Rüstungskontrolle in diesem Bereich.

Keywords: Information Warfare, Internationale Sicherheit, Abrüstung, Rüstungskontrolle, Proliferation, IT-Sicherheit, Kompromittierung

1 Information Warfare

In den vergangenen 20 Jahren haben sich die zu Beginn divergenten Ansätze zu Information Warfare in Staaten wie der VR China [Hag14], Russland [Dar2014] und den USA inhaltlich deutlich angenähert. Information Warfare ist heute zu sehen als Handlungsfeld, dessen Ziel die Aufklärung und Beeinflussung gegnerischer Entscheidungsfindung ist und explizit militärische Ziele ebenso wie Medien und nicht-staatliche Organisationen einschließt. Die technischen und nichttechnischen Mittel des Information Warfare umfassen jede Art psychologischer und elektronischer Kriegführung, den Einsatz digitaler Manipulationswerkzeuge für Aufklärung und Sabotage bis hin zur Destruktion mit konventionellen Waffen [DoD10], [DoD13]. Die verschiedenen Akteure weltweit setzen unterschiedliche Schwerpunkte auf Technik und Organisation, Psychologie und Medien. Dennoch lassen sich Information Warfare-Strategien grundsätzlich und für alle Protagonisten klar strukturieren in eine nach Einsatzmitteln und Konfliktintensität abgestufte Hierarchie aufeinander aufbauender Aktionsfelder [Ruh05].

Die durch Snowden ermöglichten Einblicke erlauben ein zutreffenderes Bild der IT-Sicherheitslage, des Standes von Information Warfare Operationen und der Akteure in den USA, Europa und in Drittstaaten [RuBe14]. Dabei ist weder eine klare Trennlinie

[1] FH Brandenburg, Fachbereich Wirtschaft, Magdeburger Straße 50, 14770 Brandenburg ruhmann@fh-brandenburg.de

zwischen geheimdienstlichen und militärischen Akteuren zu erkennen, noch ist eine solche beabsichtigt. Dieser fließende Übergang zwischen verschiedenen politischen und militärischen Aktionsfeldern hat Folgen für die Einhegung im internationalen Recht.

Als – wegen des besonderen Umstands verfügbarer Dokumente hier hervorgehobene – Beispiele seien die USA und Deutschland angeführt. In den USA wurden die bis 2008 zersplitterten militärischen und geheimdienstlichen Zuständigkeiten für Information Warfare bei der National Security Agency (NSA) zusammengeführt, die ihrer derzeit gültigen Errichtungsanordnung zufolge eine „Kampfunterstützungseinheit" des U.S. Department of Defense ist [DoDD10]. Der Leiter der NSA ist zugleich Oberkommandierender des U.S. Cyber Commands. Zur Intensivierung der Information Warfare-Kapazitäten und Stärkung des Cyber Commands kündigte U.S.-Verteidigungsminister Carter im April 2015 den Aufbau einer „Cyber Mission Force" aus zusätzlichen 6.200 Militärs, Zivilisten und kommerziellen Auftragnehmern an [DoD15]. Analog zum U.S. Cyber Command hat die Bundeswehr 2002 im „Kommando Strategische Aufklärung" (KSA) jede Form der Aufklärung, die elektronische und psychologische Kriegsführung zusammengezogen. Auf Computerangriffe spezialisiert ist im KSA die 2009 gegründete Abteilung „Informations- und Computernetzwerkoperationen" [BW10][2].

Die zugänglich gemachten Haushaltszahlten der NSA zeigen, dass für das Jahr 2013 allein für die Entwicklung von Schadsoftware die Summe von 625 Mio. Dollar beantragt wurden, weitere 10 Mrd. Dollar für das gemeinsame kryptologische Programm der NSA und verbündeter Dienste. Zum Vergleich: Der gesamten, dem U.S.-Justizministerium zugeordneten zivilen Strafverfolgung von Computerdelikten in den USA stand 2014 ein Budget von 722 Mio. Dollar zur Verfügung einschließlich der Mittel für den Schutz der eigenen IT-Systeme[Moo14]. Der Mittelansatz der NSA für Schadsoftwareentwicklung ist nicht ungewöhnlich. Anhörungen des U.S.-Kongresses 2007 zufolge wandte die NSA auch in den Jahren 2005 bis 2007 annähernd zwei Milliarden US-Dollar auf für die Entwicklung der bis heute eingesetzten Systeme „Turmoil" und „Turbulence" die der massiven Datensammlung, Manipulation und Kontrolle von Internet-Knotenpunkten und Computersabotage durch selektive Modifikation von Datenpakten dienen[3].

In der Leistung breiter ist das NSA-System „XKeyScore" [Gre13], [Lis13], das auch Bundesnachrichtendienst und Bundesamt für Verfassungsschutz eingesetzt haben [Spi13]. XKeyScore wurde bekannt als Werkzeug der NSA für die Erhebung, Analyse und Strukturierung von Inhalten und Metadaten aller Arten von Kommunikationsvorgängen. Während hierzulande über die Erkenntnisse aus der automatisierten Metadatenanalyse von Kommunikationsnetzen berichtet wurde, wies ein Bericht für das EU-Parlament kurz nach Beginn der Enthüllungen schon auf die Funktionen von XKeyScore als automatisiertes Angriffswerkzeug hin [EP13].

[2] So die Antwort der Bundesregierung zur Abteilung „Computernetzwerkoperationen" in der Bundestags-Drucksache 18/3963

[3] Turmoil und Turbulence werden im Übrigen in den NSA-Dokumenten zu XKeyScore zu einem Vergleich heran gezogen, so S. 8 der vom Guardian dokumentiertem XKeySore-Präsentation der NSA, http://www.documentcloud.org/documents/743252-nsa-pdfs-redacted-ed.html

Zu den Fähigkeiten der NSA gehört die Entschlüsselung von Kommunikation in VPN-Netzen über Router verschiedener Hersteller [App14]. Angestrebt wurde vor Jahren die Online-Entschlüsselung von SSL-Verschlüsselung, IPSec, SSH, und verschlüsselten Chats und VOIP-Verbindungen. XKeyScore leistet heute nicht nur die Suche nach verschlüsselter Datenkommunikation in Zielgebieten, sondern entschlüsselt diese Kommunikation bei bestimmten Verfahren, analysiert die kommunizierenden IT-Systeme automatisch auf Schwachstellen und schleust bei Bedarf - ebenfalls automatisiert - Schadcode ein [Lis13]. Für die Schwachstellenanalyse der Zielsysteme wird unter anderem auf abgefangene Software-Absturzberichte sowie auf eine Schwachstellendatenbank zurückgegriffen, die seit den 1980er Jahren gepflegt wird [Ber97]. Sofern sich IT-Systeme automatisierten Verfahren widersetzen oder nicht an ein Netzwerk angeschlossen sind, wird die NSA-Abteilung für „Tailored Access Operations“ (TAO) eingeschaltet, die spezifische Angriffe entwickelt, durchführt und sich bei Bedarf auch den „physischen Zugang“ zu Zielsystemen verschafft[4]. Im TAO arbeiten den Angaben zufolge mit über 900 Hacker mehr Personen, als das für die Bekämpfung von schwerwiegenden Computerdelikten zuständige FBI an Agenten für die Aufklärung von Computerdelikten bis 2015 landesweit erst zu erreichen hofft – von den geplanten 750 Agenten sind 152 erst noch einzustellen [Man14].

Die bekannt gewordenen Ziele dieser Manipulationen sind neben staatlichen Einrichtungen diverse Gegner in verschiedenen Ländern, Organisationen wie etwa die UNO, die EU-Kommission oder Personen wie die Bundeskanzlerin oder die brasilianische Staatspräsidentin. Festzuhalten ist daher, dass sich diese Art des Information Warfare gleichermaßen gegen Freund und Feind richtet. Und genau das ist seit Jahren auch der Kenntnisstand deutscher IT-Sicherheitseinrichtungen. Das bei Angriffen auf IT-Systeme der Bundeswehr eingesetzte CERT der Bundeswehr beantwortet die Frage „Wer bedroht und eigentlich?” schon seit Jahren nicht nur mit den üblichen Hackern, sondern auch mit „Traditionellen Geheimdiensten (Freund und Feind)”[5].

Auch zu den eingesetzten Werkzeugen haben Analysen der letzten Zeit zu einer verbesserten Datenlage geführt. Nachdem bei der Identifikation der Stuxnet-Schadsoftware anhand von Indizien nur gemutmaßt werden konnte, wurde 2012 publik, die gegen die iranische Uranaufbereitung gerichtete Schadsoftware sei von Diensten aus den USA und Israel verbreitet worden [San12]. Vergleichende Codeanalysen von Stuxnet und anderen Schadsoftware-Paketen führten zur Identifikation mehrerer Familien von Trojanern und Hinweisen, wonach fast ein halbes Dutzend Schadsoftwarepakete mit hinreichender Wahrscheinlichkeit von denselben staatlichen Urhebern wie Stuxnet stammen könnten [Gos12] [Kas12].

Mittlerweile lässt sich gut dokumentieren, dass die NSA die ressourcenstärkste Hackertruppe der Welt ist, die einen Cyberkrieg gegen Freund und Feind führt [Ruh14]. Doch ist hier immer wieder darauf hinzuweisen, dass diese Aussage nur ein Ergebnis einer

[4] Auch über diesen Modus Operandi wurde bereits in den 1980er Jahren berichtet [Pet89]
[5] Siehe schon2009 S. 3 des Folienvortrages des CERT Bw auf der AFCEA 2009: http://www.afcea.de/fileadmin/downloads/Young_AFCEAns_Meetings/20090216%20Wildstacke.pdf

ungewöhnlich guten Quellenlage ist, und Nachrichtendienste und Militärs anderer Länder dem im Rahmen ihrer Möglichkeiten nacheifern.

So ist die im Frühjahr 2015 in ihren Aktionen analysierte „Great Cannon“ ein Angriffswerkzeug, das Datenverkehre in die VR China gezielt kapert, die Inhalte der übertragenen Daten verändert und zusätzlich IT-Systeme außerhalb Chinas so manipuliert, dass diese in eine massive verteilte Denial-of-Service-Attacke (DDoS-Angriff) eingebunden werden [Mar15]. Die Eigenschaften dieses Systems und seines Umfeldes lassen auf staatliche Akteure als Urheber schließen.

Vorfälle in der Zeit von 2008 bis 2015 wie die Manipulationen von IT-Systemen vor dem Konflikt zwischen Russland und Georgien, im Kontext einer vornehmlich propagandistischen Auseinandersetzung zwischen Russland und Estland, das Eindringen in die IT-Systeme von Sony sowie die zeitweilige Übernahme der IT des TV-Senders TV 5 Monde unterstreichen, dass die Nutzung von Manipulations- und Sabotagewerkzeugen zum Alltag bei der Austragung von Konflikten nahezu jeder Art gehört. Spätestens seit Stuxnet hat Information Warfare auch die Schwelle zur Verursachung physischer Schäden überschritten.

2 Umfassende Kompromittierung von IT-Systemen

Wer sich mit Information Warfare als Facette von IT-Sicherheit beschäftigt hat, wird die Frage stellen, was neu an dieser Lagebeschreibung ist. Die Supermächte USA und Sowjetunion drangen bereits Mitte der 1980er Jahre wechselseitig in IT-Systeme ein, um Daten zu entwenden und Computer zu sabotieren [Pet89] [Sen86]. Für Propagandazwecke gekaperte und manipulierte Web-Services waren Mitte der 1990er Jahre zum Werkzeug der Konfliktaustragung geworden. Die systematische Erhebung und Analyse von Sicherheitslücken in IT-Systemen wird seit den 1990er Jahren betrieben, um Lücken bei Bedarf zu nutzen[6]. Neu sind an der heutigen Situation vor allem zwei Faktoren:

1) Die seit 20 Jahren systematisch betriebene Entwicklung von Zugangs- und Manipulationsmöglichkeiten zu IT-Systemen, die Verfügbarkeit von Mitteln und vor allem von Ressourcen auf Angreiferseite sowie die zunehmende Vernetzung aufseiten potentieller Ziele waren Voraussetzung für die heute erkennbare Ausweitung der ehemals primär auf staatliche Ziele (Militär, Geheimdienste, Verwaltung) gerichteten Angriffe auf nun nahezu beliebige Ziele. Ein Beispiel dafür ist das Unternehmen Sony oder Sekundärziele – wie etwa der Angriff auf die Individualkommunikation von Systemadministratoren der Belgacom, um auf diesem Weg Zugang zu den Systemen dieses Providers und dessen Kunden, der EU-Kommission, zu erhalten.

[6] So existiert die noch heute von der NSA genutzte Schwachstellen-Datenbank ”Constant Web” seit den 1980er Jahren, als sie als Werkzeug für die Elektronische Kriegsführung aufgebaut wurde: Air Intelligence Agency – History Retrospective 1980s; http://usafss_boltinghouse.tripod.com/usafss_history/1980s.html

2) Die – als spiegelbildlicher Effekt - heute starke Abhängigkeit der Zivilgesellschaft von funktionierenden IT-Systemen, die oft gravierenden Folgen eines isolierten Ausfalls und die Anfälligkeit der Systeme gegenüber Manipulationen, die in hochgradig interdependenten Systemverbünden größere, kaskadierende Schäden schon bei Angriffen auf eigentlich unbedeutende und daher weniger gut geschützte Systeme zur Folge haben.

Die durch Snowden ermöglichten Einblicke zeigen, in welchem Ausmaß IT-Sicherheitssysteme – Verschlüsselungssysteme, digitale Signaturen, Hardware-bezogene Sicherungssysteme – kompromittiert sind. Dennoch verbleibt die Debatte um den NSA-Skandal auf der Ebene des Datenschutzes. Dabei war diese absehbare Entwicklung schon 1998 Grund für die Feststellung, dass die Abwicklung jeder Tätigkeit per Internet die Kontrolle der Telekommunikation zum ”strategischen Grundrecht” mache, und die Freiheit der Informationsgesellschaft davon abhänge, welches Ausmaß diese Überwachung haben werde [Kip98].

Auch wenn Datenschutz heute für manche überholt scheint, so kann dies bei der Manipulation kritischer Infrastrukturen nicht gelten: Es ist keine Option, die Sicherheit der IT-Systeme zur Steuerung etwa der Energieversorgung oder von Chemiewerken in bewohnten Umgebungen dem Belieben von Militärs und Geheimdiensten zu überlassen. Es ist auch keine Option zur Reaktion auf solche Risiken, Eingriffe in IT-Infrastrukturen durch einen Rüstungswettlauf klassischer Art zu lösen. Dieser würde bedeuten, alltägliche IT-Systeme sicherheitstechnisch zu „verbunkern” und so ihre Nutzbarkeit stark zu reduzieren, ohne damit aber die Sicherheitslücken in den IT-Systemen systematisch zu schließen.

Die Snowden-Dokumente zeigen über die NSA und ebenso deren Gegner, dass Überwachung nur der Ausgangspunkt für die Manipulation von IT-Systemen ist und, dass grundlegende Sicherheitsmechanismen der IT wie Software-Zertifikate, Verschlüsselungs- und Authentisierungsmechanismen kompromittiert sind. Das Ausmaß dieser Kompromittierung ist bis heute nicht ernsthaft bemessen.

Die durch Snowden erlaubten Einblicke in eine durch Militärs und Geheimdienste fundamental korrumpierte IT-Sicherheitsarchitektur machen es zu einem Irrglauben, allein technische Lösungen würden Abhilfe schaffen können und einer von IT-Infrastrukturen abhängigen Gesellschaft ein erträgliches Maß an Sicherheit und Zuverlässigkeit bringen. Die heutigen Konzepte der IT-Sicherheit sind nicht annähernd ausreichend vor dem Hintergrund der Eingriffsmöglichkeiten und Ressourcen, die Militärs und Geheimdiensten zur Verfügung gestellt werden und weit jenseits jeder anderen Gruppierung liegen. Ausgangspunkt für neue Konzepte muss daher die Berücksichtigung einer militärischen Ratio von Eingriff und Zerstörung sein, die im Zivilleben kein Pendant kennt.

3 Sicherheitspolitische Bewertung

Die Auswirkungen auf die Verfügbarkeit und Integrität von IT-Systemen ist zwar die primäre, aber nicht die einzige Folge dieser Entwicklung von Information Warfare. Cyberattacken haben nicht nur ein Eskalationspotential in IT-Umgebungen, sondern vor allem auch sicherheitspolitische und militärische Folgen in der realen Welt.

So gehört es seit 2011 zur Doktrin der USA, Cyberangriffe als kriegerischen Akt anzusehen und darauf auch mit konventionellen Waffen zu reagieren [DoD11]. Ob und wann es sich bei Information Warfare-Operationen um kriegerische Aktionen handelt, arbeitete eine von der NATO 2011 berufene Expertengruppe aus der Analyse und Bewertung internationalen Rechts auf, um einen Rahmen für den Umgang mit Information Warfare-Aktionen zu finden. Das Ergebnis, das so genannte „Tallinn-Manual", fokussiert auf staatliche Akteure, betrachtet aber ein Spektrum von Spionage über Cyber-Söldner bis zu militärischen Aktionen und daher auch zwischenstaatliche Rechtsakte jenseits des klassischen Kriegsrechts. Die wesentlichen Ergebnisse sind, dass Cyber-Angriffe,

- die von einem Staatsgebiet ausgehen, aber nicht von staatlichen Stellen orchestriert werden, einen Bruch internationalen Rechts darstellen, wenn sie Schaden anrichten und nicht von den nationalen Strafverfolgern unterbunden werden.
- die von staatlichen Stellen - wie etwa der NSA oder dem GCHQ - verübt werden, nach internationalem Recht als Kriegshandlungen gewertet werden können. Die NATO-Rechtsexperten halten es aufseiten der angegriffenen Staaten für gerechtfertigt, gleichwertige Gegenmaßnahmen zu ergreifen [Sch13, S. 29ff].
- die die Wirkung von Militärschlägen in der Zahl der Opfer, oder dem Grad der Zerstörungen erreichen, den NATO-Experten zufolge auch militärische Reaktionen erlauben[Sch13, S. 63ff].

Wenn Cyberattacken Kriegshandlungen sein und konventionelle militärische Reaktionen auslösen können, folgt die Frage, wann dadurch das Kriegsvölkerrecht verletzt wird und Kriegsverbrechen begangen werden - eine Frage, mit der sich auch das am Tallinn Manual beteiligte International Rote Kreuz (IKRK) auseinandergesetzt hat. Die vom IKRK kommunizierte Position ist, dass Cyberangriffe die Zivilbevölkerung und zivile Infrastrukturen ausnehmen und unvermeidbare Schäden begrenzt bleiben sollten oder auf Cyberangriffe verzichtet werden solle [ICRC13]. [Lin12] wiederum geht bei Konflikten von organisierten Gruppen aus und sieht einen bewaffneten Konflikt zusätzlich auch als die Absicht, Schaden zu verursachen. Auf netzbasierte Aktionen übertragen schlussfolgert Lin, dass das Verursachen von schweren Schäden durch Cyberattacken eine Angriffshandlung nach dem Kriegsvölkerrecht darstelle.

Nach diesen Lesarten ist mit den physischen Schäden durch Stuxnet die Schwelle zur bewaffneten Auseinandersetzung und zum Recht auf einen Gegenschlag erreicht. Zudem ist dabei am deutlichsten die Beteiligung regulärer Kombattanten erkennbar und eine kriegerische Handlung mit einem militärischen Ziel. Mit Stuxnet oder dem vom BSI

jüngst beschriebenen Angriff auf einen Hochofen in Deutschland sind die Grenzen sowohl zu physischen Schäden als auch – nach dem Kriegsvölkerrecht rechtswidrigen - Angriffen auf die Zivilbevölkerung überschritten.

Wichtig ist, auch hier die Perspektive über die USA hinaus auszuweiten. Auf die parlamentarische Frage, ob die in der Bundeswehr zur Ausübung von Cyberattacken eingerichtete Gruppe „Computer Netzwerk Operationen" innerhalb des Kommandos Strategische Aufklärung (KSA) auch dafür trainiere, in IT-Netzwerke ziviler kritischer Infrastrukturen einzudringen, diese zu manipulieren und zu schädigen, antwortete der Parlamentarische Staatssekretär bei der Bundesministerin der Verteidigung:

> „Aus rechtlicher Sicht gelten auch elektronische Systeme während eines bewaffneten Konflikts als militärische Ziele, wenn sie entsprechend der Definition des humanitären Völkerrechts aufgrund ihrer Beschaffenheit, ihres Standorts, ihrer Zweckbestimmung oder ihrer Verwendung wirksam zu militärischen Handlungen beitragen und deren gänzliche oder teilweise Zerstörung oder Neutralisierung einen eindeutigen militärischen Vorteil darstellt. […] Das humanitäre Völkerrecht bestimmt keinen absoluten Schutz von Energieversorgungseinrichtungen, des Transportwesens oder der Telekommunikation. Ob ein elektronisches System ein ziviles Objekt oder aber ein militärisches Ziel darstellt, kann nur unter Berücksichtigung aller Umstände des konkreten Einzelfalls bestimmt werden."[7]

Aus Perspektive der IT-Sicherheit geht es hier darum, IT-Systeme in zivilen Umgebungen, die von Angreifern kompromittiert sind und für Angriffe genutzt werden – also als Ausgangspunkte für Attacken missbraucht werden -, ebenso anzugreifen, wie etwa in konventionellen Konflikten ein ziviles Gebäude, das Operationsbasis von gegnerischen Militärs ist. Der wesentliche, in der parlamentarischen Antwort jedoch nicht geklärte Unterschied besteht allerdings darin, legitime einzelne virtuelle oder realweltliche Ziele auszuschalten oder – der Frage gemäß - zivile Infrastrukturen als Antwort auf Cyberattacken lahmzulegen. Auch hier gilt, dass Angriffe gegen andere Staaten in Form einer vorbereitenden Schadsoftware-Verbreitung oder von gezielten Angriffen auf Kommunikationsinfrastrukturen, Energienetze oder andere lebenswichtige zivile Infrastrukturen zu werten ist als bewaffneter Konflikt ausgetragen auf zivilen digitalen Infrastrukturen.

IT-Sicherheit ist muss sich also heute damit auseinandersetzen, dass Information Warfare ein Kriegshandeln auf IT-Infrastrukturen darstellt, das mit militärischer Logik, hoher Zerstörungsintensität und exorbitanten Ressourcen ausgetragen wird.

4 Ansätze zur Begrenzung

Die primäre Schlussfolgerung aus dem Einsatz von Information Warfare-Mitteln und deren grundsätzlich internationale Dimension einerseits und die immer noch deutlich ausbaufähigen Kooperationsansätze zur Bekämpfung von Cyberangriffen andererseits ist

[7] Plenarprotokoll der Fragestunde der 16. Sitzung des Deutschen Bundestages vom 19.02.2014, Antwort auf Frage Nr. 8, S. 1165f

in technischer Hinsicht die Reduktion von Angriffspfaden durch robustere IT-Systeme. Die offensichtlichen Asymmetrien zwischen den Investitionen zugunsten der Kompromittierung von IT-Systemen und zum Nachteil der Entwicklung von Schutztechniken verdeutlichen, dass nur zusätzliche Anstrengungen, Information Warfare auf internationaler Ebene zu begegnen, eine Chance bieten, die Funktionsfähigkeit der sich heute global auf funktionierende IT-Systeme abstützenden Gesellschaft zu erhalten.

Ein Ansatz zur Begrenzung eines Rüstungswettlaufs beim Einsatz von IT-gestützten Manipulationswerkzeugen war auf nationaler Ebene eine bereits 1994 vom Deutschen Bundestag in Auftrag gegebene Analyse [Kli95], deren Schlussfolgerung war, dass „der Bereich neuer Schutzkonzepte entsprechende Anstrengungen erforderlich machen“ werde [BT95], was bislang kaum verfolgt wurde [Ben02] [Alt07] [Neu14]. Der heute erreichte Stand bei Information Warfare macht stärkere Anstrengungen dringlich.

Doch statt einer Eindämmung militärisch-geheimdienstlicher Aktivitäten haben über 100 Staaten Information Warfare-Kapazitäten aufgebaut, die immer zuerst defensiv ausgerichtet sind und über die Zeit dann offensive Aspekte umfassen. Das UN-Institut für Abrüstungsforschung UNIDIR hatte 2013 bereits in Medienverlautbarungen 41 Staaten ausgemacht, die militärische Cyberaktivitäten verfolgen [UNID13]. Die Folge ist eine weiter steigende digitale und reale Eskalationsgefahr.

4.1 Zwischenstaatliche Kooperation

Cyberangriffe werden über IT-Systeme in verschiedenen Rechtsräumen abgewickelt, um die Verfolgung jenseits nationaler Grenzen zu erschweren oder zu verhindern. Jede Reaktion auf Cyberangriffe wird deshalb die Notwendigkeit einer Kooperation mit Akteuren im internationalen Umfeld entwickeln.

Zu Unterstützung der schnellen Kooperation in der Strafverfolgung wurde 2001 das Cybercrime Abkommen durch den Europarat verabschiedet und von inzwischen über 40 Staaten unterzeichnet. Nach Artikel 27 der Konvention ist allerdings genau der Fall von der Kooperation ausgenommen, bei dem es um Vorfälle geheimdienstlicher Natur oder von nationaler Sicherheit geht. Der ENISA-Präsident Helmbrecht wies zusätzlich auf Schwierigkeiten einer Kooperation staatlicher Stellen hin, wenn die für IT-Sicherheit zuständigen Behörden im zivilen und militärischen Bereich angesiedelt sind. So würde eine zivile Behörde - wie etwa das Bundesamt für Sicherheit in der Informationstechnik (BSI) – etwa mit der in Frankreich im Präsidialamt und im Umfeld der Zuständigkeiten für die Geheimdienste angesiedelten Agence Nationale de la Sécurité des Systèmes d’Information (ANSSI) zwar Informationen austauschen, von dort oder gar vom britischen Geheimdienst-Pendant GCHQ aber nicht vollständig mit relevanten Informationen in dem Fall versorgt, wenn dieser als nachrichtendienstlich relevant bewertet würde.

Schlussfolgerung für die zivile internationale Kooperation zur Bekämpfung von Cyberangriffen ist zwar das Vorhandensein von Kooperationsstrukturen, verbunden aber mit großen Umsetzungsschwierigkeiten in der Praxis.

4.2 Proliferationskontrollen

Als Surrogat für eine bessere aktive Kooperation wurde versucht, einen international abgestimmten, aber in der Umsetzung national zu verfolgenden Weg zu wählen, der an der Technik ansetzt. Um Ländern des COMECON während des Kalten Krieges den Zugang zu spezifischen Hochtechnologieprodukten zu erschweren, nahm 1950 das "Koordinierungskomitee für mehrseitige Ausfuhrkontrollen" - kurz COCOM - seine Arbeit auf zur Abstimmung der Technologieexportgesetze der NATO-Staaten und weiterer Verbündeter. In den 1970er Jahren begannen die USA, Exportkontrollen von Dual-Use-Gütern extensiv auf die Informationstechnik auszuweiten [Ruh03]. Nach dem Ende von COCOM 1994 wurde die Kontrolle von Rüstungsexporten im Wassenaar-Abkommen weitergeführt, dem auch Russland angehört. Verboten sind nach danach auch heute beispielsweise Exporte in bestimmte Länder von Supercomputern, Simulations-Software im Kontext der Konstruktion von Massenvernichtungswaffen, oder Ver- und Entschlüsselungstechnologie. Auf Betreiben der britischen und französischen Seite und als Reaktion auf die Aufdeckung von entsprechenden Exporten in den arabischen Raum wurde Ende 2013 das Wassenaar-Regularium weiter verschärft. Sofern man Cyberwaffen-Technologie als (vornehmlich Software-) Systeme zur Manipulation und Sabotage von IT-Systemen beschreibt, dann ist das Wassenaar-Reglement also auch als Ansatz zur Proliferationskontrolle zu sehen.

Für jede bisherige Kontrolle von Rüstungstechnologie war es nicht ausreichend, nur auf die Non-Proliferation von Technologie zu setzen. Der einfache Verbreitungsweg von Cyberwaffen – meist intangible Software – erschwert zusätzlich die Kontrolle entsprechender Technologien.

4.3 Neue Erkenntnisse systematisieren - neue Abrüstungsansätze entwickeln

Bisherige Rüstungskontroll-Ansätze setzten vornehmlich auf der Ebene von „Cyberwaffen“ an. Die durch die enthüllten NSA-Dokumente ermöglichten Einblicke in die Methoden und Verfahren der NSA, ihrer Partnerdienste und deren Erkenntnisse über gegnerische Dienste haben den Nachbau von Spionagewerkzeugen wie etwa die Klasse der USB-Schadsoftware und andere Werkzeuge beschleunigt[8]. Solches Detailwissen ist auch ein Ausgangspunkt für die Entwicklung neuer Maßnahmen für Abrüstung und Rüstungskontrolle. Entscheidend ist dabei, dass zusätzlich zur Diskussion über „Cyberwaffen“ nun die Überlegungen über eine Herangehensweise bei Abrüstungsansätzen unter anderem auch auf Infrastrukturen, auf Ressourcen und Personalstärke, auf technische Werkzeugen und viele andere Aspekten ausgeweitet werden können. Diese Ansätze können hier nicht erschöpfend behandelt werden, doch sollen im Folgenden Anregungen für weitere Arbeiten auf diesem Forschungsfeld gegeben werden.

Infrastrukturen

[8] Hacker bauen NSA-Backdoors nach; Golem.de, 20.06.2014. Zur Sammlung der Ergebnisse wurde die Webseite http://www.nsaplayset.org/ eingerichtet

Im ersten Teil dieses Beitrages wurde dargelegt, dass die Kommunikationsüberwachung zugleich Ausgangspunkt und Werkzeug für militärisch-geheimdienstliche „Cyber"-Angriffsoperationen ist. Diverse Werkzeuge stellen sich als integrierte Überwachungs- und Angriffssysteme dar, die spezifische Infrastrukturen mit physisch nachprüfbaren Eigenschaften benötigen, die eine Kontrolle auch in einem Rüstungskontrollregime möglich machen.

Infrastrukturen für Cyberangriffe sind auffallend und für die zivile Welt ungebräuchlich. Die Operationen der „Great Cannon" chinesischer Akteure waren im Internet nicht zu übersehen. Die Leitungskapazitäten der heute als einer von weltweit drei Übergangsknoten für spezielle NSA-Angriffsoperationen bekannten Station in Vaihingen bei Stuttgart[9] waren für IT-Fachleute schon vor Jahren nicht zu übersehen, aber mangels Wissen über deren Zwecke nicht bewertbar. Der Einsatz von Werkzeugen der „Quantum"-Technologie für automatisierte Man-in-the-Middle-Attacken auf durchlaufenden Datenverkehr wiederum setzt voraus, leistungsfähige IT in die Nähe der Operationsziele zu platzieren. Für die Zugänge des GCHQ zu Unterseekabeln, des BND bei der Überwachung des Satelliten-Datenverkehrs und andere Zugänge zu bestehenden IT-Systemen und Infrastrukturen gilt, dass die zivilen Infrastrukturen den Ort und die Zugangsmöglichkeiten vorgeben, um dort Information Warfare-Werkzeuge anzubinden, die weder örtlich beliebig noch beliebig klandestin positionierbar sind. Das Ableiten von Datenverkehr und das Injizieren von Schadcode setzt entweder Rechenleistung oder angemessen dimensionierte zusätzliche Datenübertragungskapazität voraus, die nicht unentdeckt bleiben kann. Solche Infrastrukturen sind zudem an spezifische Gegebenheiten vielfach zentraler Knotenpunkte angepasst und nur schwer zu ersetzen. Es fehlten bisher aber die Aufmerksamkeit und der Wille, solche Auffälligkeiten zu erheben, zu systematisieren und für Zwecke der Rüstungskontrolle nutzbar zu machen. Eine systematische Erhebung von Infrastrukturen würde auf dieser Ebene die Basis für Verifikationsansätze bei Rüstungskontrollvereinbarungen schaffen, indem etwa Backbones auf Ableitungen und Angriffe gezielt überwacht und bei vor-Ort-Inspektionen geprüft würden.

IT-Kooperationsstrukturen für Rüstungskontrolle nutzen und ausweiten

Eines der wirkungsvollsten Instrumente gegen Angriffe ist der Austausch zwischen Computer Emergency Response Teams – CERTs. Eine internationale Kooperation einer Gruppe interessierter Staaten, aber auch von Organisationen und betroffenen Unternehmen – vorzugsweise in Verbindung mit dem Aufbau einer internationalen Organisation – könnte dazu führen, Daten zu Angriffen aufzuzeichnen und zu analysieren, und somit Verifikationsansätze liefern.

Für Verifikationsbemühungen ist es ebenso wichtig, die Forensik bei Cyberangriffen zu intensivieren und dies nicht allein den Mitteln und Möglichkeiten privater IT-

[9] Die beiden anderen Stationen sind angesiedelt in Wiesbaden und in Yokota, Japan, so die Darstellung über das NSA-Programm TUTELAGE zur Nutzung von Ergebnissen der Computersabotage durch gegnerische Dienste (Ergebnisse von dritter und vierter Seite), zitiert nach der NSA-Präsentation in Spiegel Online: http://www.spiegel.de/media/media-35685.pdf , Folie 5

Sicherheitsunternehmen zu überlassen. Der Grund dafür ist weniger deren Glaubwürdigkeit, sondern die Dauer der Untersuchungen: So wurde die Stuxnet-Schadsoftware von IT-Sicherheitsunternehmen der Öffentlichkeit präsentiert. Die Analyse von Details und daraus ableitbare weitere Schlussfolgerungen aber ist mangels Ressourcen keineswegs abgeschlossen und führt bis heute zu weiteren Ergebnissen. Für die Begrenzung einer Eskalation im konkreten Krisenfall kommen solche Analysen jedoch zu spät. Ähnlich der Befugnisse und Aufgaben der zur Verifikation bei Bio- bzw. Chemiewaffen gegründeten Organisationen wäre eine internationale IT-Forensik-Einrichtung ein wichtiger Schritt zur Entschärfung von Cyberattacken.

Die Aktivitäten der NSA zur Gewinnung von Daten gegnerischer Cyberangreifer und die genutzten Methoden und Werkzeuge zeigen darüber hinaus, dass Cyberangreifer regelmäßig Einblicke in die Strukturen, Mittel und Ressourcen ihrer Gegner haben. Den NSA-Dokumenten zufolge gibt es eine eigene Form von Angriffen, um die Ergebnisse von Cyberangreifern der Gegenseite, von dritter oder gar vierter Seite zu nutzen und diese zu manipulieren[10]. Wenn die Dienste ihre Erkenntnisse zur Gegenspionage und –sabotage nutzen, dann gibt es keinen sachlichen Grund, auf eine auf konkrete Einrichtungen bezogene Form der Überwachung von Cyber-Rüstung zu verzichten. Bei der Kontrolle von ABC-Waffen wird heute auf geheimdienstliche Erkenntnisse zurück gegriffen, um die Verbreitung zu begrenzen. Offensichtlich gilt im Information Warfare ähnliches. Es ist nicht begründbar, warum solche Erkenntnisse allein für die Auseinandersetzungen unter Geheimdiensten reserviert sein sollten. Für die Rüstungskontrolle von Information Warfare ließe sich daher das Grundprinzip der Nutzung geheimdienstlicher Erkenntnisse ebenso anwenden.

Begrenzung von Ressourcen und Truppenstärken

Das Erheben und Zählen von Panzern, Soldaten, Kasernen und Infrastrukturen sowie die zur Kontrolle nötigen Rechte für Kontrollen vor Ort sind ein in der Rüstungskontrolle lange geübtes Verfahren, um Abrüstungsvereinbarungen zu verifizieren. Die Analyse von Ressourcen für Information Warfare und deren Kontrolle könnten einen neuen und zusätzlichen Weg darstellen, um die Rüstung im „Cyberspace" zu begrenzen.

Information Warfare setzt zunehmend auf speziell entwickelte automatisierte Werkzeuge. Zur Entwicklung, Ausübung und Kontrolle von Angriffsaktionen werden Spezialisten benötigt, die nicht in beliebiger Zahl verfügbar sind. Techniken wie die „Quantum"-Werkzeuge lassen sich zudem nur mit spezifischen Werkzeugen und Spezialisten und nicht durch ein Heer von „Cyber-Fußsoldaten" realisieren. Es fehlt derzeit für Rüstungskontrollansätze jedoch an einer grundlegenden Analyse von materiellen und personellen Ressourcen für spezifische Cyber-Angriffsformen und deren Einsatz. Grundsätzlich ist die Idee, Information Warfare-Aktionsformen zu verstehen und die Akteure und ihre Ressourcen zahlenmäßig zu begrenzen, aber ein wichtiger Ansatz, um die Handlungsspielräume von Cyberangreifern zu begrenzen.

[10] Das NSA-Programm mit der Bezeichnung „Transgression" zielt ab auf Ergebnisse von dritter, vierter und fünfter Seite, so [App15]

Ein ganz wesentlicher Teil dieser Begrenzung ist die systematische Analyse der Aufwände für Information Warfare. Aus den Snowden-Dokumenten, Daten aus den USA, Deutschland und anderen Ländern geht hervor, dass für Entwicklungen zur Aushebelung von IT-Sicherheitsmechanismen bis zu zehnmal mehr Ressourcen aufgewandt werden, als für die Erforschung und Entwicklung von IT-Schutzmechanismen. Solche Relationen werden – anders als Daten zu konventionellen Rüstungsaufwendungen, wie sie etwa durch das SIPRI erhoben werden – bisher nicht systematisch erhoben und ausgewertet. Erst eine solche Datenanalyse erlaubt es der Öffentlichkeit, Entscheidungen von Regierungen und Parlamenten zu hinterfragen und weiter gehende Rüstungkontroll-Anstrengungen zu entwickeln. Denn IT-Sicherheit wird nicht durch die Aufrüstung von Information Warfare-Akteuren herzustellen sein, sondern durch Rüstungskontrolle und wesentlich verstärkte Forschungs- und Entwicklungsanstrengungen für robuste Systeme, und grundlegend neue Konzepte und Techniken für sichere Systeme in unsicheren Umgebungen.

5 Fazit

Ein besonderer Wert der NSA-Dokumente liegt in der Möglichkeit tiefer Einblicke in ein bisher im Verborgenen betriebenes Aktionsfeld von Militärs und Geheimdiensten und in einem besseren Verständnis der Techniken und Strukturen des Information Warfare. Information Warfare erweist sich als Handlungsfeld, auf dem gezielt nach technischen Mitteln und organisatorischen Wegen gesucht werden kann und muss, um Muster von Cyberangreifern, ihre Spuren und Infrastrukturen zu identifizieren und zu analysieren. Bisher nicht geleistet ist die Analyse der Anwendbarkeit klassischer Rüstungskontrollmechanismen auf Information Warfare vor dem Hintergrund der heute verfügbaren Fakten. Es ist nicht länger abwegig, beispielsweise strukturelle Aspekte der Überwachung von Alltagschemie im Rahmen der Kontrolle des Chemiewaffenprotokolls oder bei der Verifikation des Atomwaffen-Teststopps zu übertragen auf die Gegebenheiten des Information Warfare.

„Cyber"-Rüstungskontrolle ist eine alte Aufgabe mit neuen Perspektiven. Die heutigen Erkenntnisse bieten zahlreiche Ansätze für neue Herangehensweisen, für die substantielle Forschungsarbeiten zu leisten sind. Die Frage ist allein, ob dieser Weg aus Klugheit und Einsicht beschritten wird oder erst als Ergebnis schwerer Schäden durch Cyberattacken.

Literaturverzeichnis

[Alt07] Altmann, Jürgen; Bernhardt, Ute; Nixdorff, Kathryn; Ruhmann, Ingo; Wöhrle, Dieter: Naturwissenschaft – Rüstung – Frieden, Wiesbaden, 2007

[App14] Applebaum, Jacob u.a.: Was die NSA knacken kann - und was nicht, Spiegel Online, 29.12.2014, http://www.spiegel.de/netzwelt/netzpolitik/snowden-dokument-so-

unterminiert-die-nsa-die-sicherheit-des-internets-a-1010588.html

[App15] Appelbaum, Jacob u.a.: Die NSA rüstet zum Cyber-Feldzug, 18.01.2015, http://www.spiegel.de/netzwelt/netzpolitik/snowden-dokumente-wie-die-nsa-digitale-kriege-vorbereitet-a-1013521.html

[Ben02] Bendrath, Ralf; Drossou, Olga; Minkwitz, Olivier: Rüstungskontrolle im Cyberspace. Perspektiven der Friedenspolitik im Zeitalter von Computerattacken. Dokumentation einer Internationalen Konferenz der Heinrich-Böll-Stiftung am 29./30. Juni 2001 in Berlin, Dokumentationen der Heinrich Böll Stiftung Nr. 20, Berlin 2002.

[Ber97] Bernhardt, Ute; Ruhmann, Ingo: Der digitale Feldherrnhügel. Military Systems – Informationstechnik für Führung und Kontrolle. Wissenschaft und Frieden, Dossier Nr. 24, Februar 1997

[BT95] Bericht des Ausschusses für Bildung, Wissenschaft, Forschung, Technologie und Technikfolgenabschätzung: Kontrollkriterien für die Bewertung und Entscheidung bezüglich neuer Technologien im Rüstungsbereich. Bt.-Drs. 13/6449

[BW10] Informationsprofis arbeiten enger zusammen; Bundeswehr-Pressemeldung vom 29.06.2010; http://www.opinfo.bundeswehr.de/portal/a/opinfo/!ut/p/c4/04_-SB8K8xLLM9MSSzPy8xBz9CP3I5EyrpHK94uyk-PyCzLy0fL3SvOLUotT4HL0qq-ACIyC_QK01NSi1KT0xK1S_IdlQEAJFZpok!/

[Dar14] Darczewska, Jolanta: The Anatomy of Russian Information Warfare. The Crimean Operation. A Case Study, Centre for Eastern Studies, Warschau, 2014

[DoD10] U.S. Department of Defense Joint Terminology for Cyberspace Operations, Washington, Nov., 2010, Washington, Nov., 2010, http://www.nsci-va.org/CyberReferenceLib/2010-11-Joint%20Terminology%20for%20Cyberspace%20Operations.pdf

[DoD13] U.S. Department of Defense: Field Manual 3-13. Inform and Influence Activities, Jan. 2013.

[DoD11] U.S. Department of Defense: Strategy for Operating in Cyberspace, Washington, July 2011; http://www.defense.gov/home/features/2011/0411_cyberstrategy /docs/DoD_Strategy_for_Operating_in_Cyberspace_July_2011.pdf

[DoD15] U.S. Department of Defense: The DoD Cyber Strategy, Washington, April 2015, http://www.defense.gov/home/features/2015/0415_cyber-strategy/Final_2015_DoD_CYBER_STRATEGY_for_web.pdf

[DoDD10] U.S. Department of Defense Directive 5100.20, National Security Agency / Central Security Service (NSA/CSS), Washington, 26.01.2010, http://www.dtic.mil/whs/directives/corres/pdf/510020p.pdf

[EP13] European Parliament: The US surveillance programmes and their impact on EU citizens' fundamental rights, PE 474.405, Brüssel, September 2013

[Gor07] Gorman, Siobhan: Costly NSA initiative has a shaky takeoff, Baltimore Sun, Feb. 11, 2007, http://articles.baltimoresun.com/2007-02-11/news/0702110034_1_turbulence-cyberspace-nsa.

[Gre13] Greenwald, Glen: XKeyscore: NSA tool collects 'nearly everything a user does on the

internet', in: The Guardian, 31.07.2013, http://www.theguardian.com/world/2013/jul/31/nsa-top-secret-program-online-data.

[Gos12] Alexaner Gostev: Kaspersky Security Bulletin 2012. Cyber Weapons, http://www.securelist.com/en/analysis/204792257/Kaspersky_Security_Bulletin_2012_Cyber_Weapons

[Hag14] Hagestad, William: Chinese Information Warfare Doctrine Development 1994 – 2014, 2014, Kindle Edition

[ICRC13] International Committee of the Red Cross: What limits does the law of war impose on cyber attacks?, 28-06-2013, https://www.icrc.org/eng/resources/documents/faq/130628-cyber-warfare-q-and-a-eng.htm

[Kas12] Kaspersky Lab Research Proves that Stuxnet and Flame Developers are Connected, June 11, 2012, http://www.kaspersky.com/about/news/virus/2012/Resource_207_Kaspersky_Lab_Research_Proves_that_Stuxnet_and_Flame_Developers_are_Connected

[Kip98] Kiper, Manuel; Ruhmann, Ingo: Von der Datenflut zur Abhörwut; in: Blätter für deutsche und internationale Politik, Nr. 3, 1998, S. 312-319

[Kli95] Klischewski, Ralf; Ruhmann, Ingo: Ansatzpunkte zur Entwicklung von Methoden für die Analyse und Bewertung militärisch relevanter Forschung und Entwicklung im Bereich Informations- und Kommunikationstechnologie; Gutachten für das Büro für Technikfolgenabschätzung des Deutschen Bundestages, Bonn, März, 1995

[Lin12] Lin, Herbert: Cyber conflict and international humanitarian law, International Review of the Red Cross, No. 886, 30.06.2012

[Lis13] Lischka, Konrad; Stöcker, Christian: NSA-System XKeyscore: Die Infrastruktur der totalen Überwachung, Spiegel Online, 31.07.2013, http://www.spiegel.de/netzwelt/netzpolitik/xkeyscore-wie-die-nsa-ueberwachung-funktioniert-a-914187.html

[Man14] Mandak, Joe: FBI cybersquad in Pittsburgh to add agents, USA Today, 17.08.2014, http://www.usatoday.com/story/money/business/2014/08/17/premier-fbi-cybersquad-in-pittsburgh-to-add-agents/14196829/

[Mar15] Marczak, Bill u.a.: China‘s Great Cannon, Toronto, 10.04.2015, https://citizenlab.org/2015/04/chinas-great-cannon/

[Moo14] Moore, Jack: Here’s Why Cybersecurity Remains a Challenge for the Justice Department; 2.12.2014, http://www.nextgov.com/cybersecurity/2014/12/heres-why-cybersecurity-remains-challenge-justice-department/100271/

[Neu14] Neuneck, Götz: Cyberwarfare – Hype oder Bedrohung? In: Ethik und Militär, Nr. 12, 2014, http://www.ethikundmilitaer.de/fileadmin/Journale/2014-12/Cyberwarfare_-_Hype_oder_Bedrohung_-_Goetz_Neuneck.pdf

[Pet89] Peterzell, Jay: Spying and Sabotage by Computer; in: Time, March 20, 1989, S. 41

[RuBe14] Ruhmann, Ingo; Bernhardt, Ute: Information Warfare und Informationsgesellschaft. Zivile und sicherheitspolitische Kosten des Informationskriegs; in: Wissenschaft und Frieden, Heft 1, 2014, Dossier Nr. 74, S. 1-16; http://wissenschaft-und-frieden.de/seite.php?dossierID=078

[Ruh03] Ruhmann, Ingo; Schulzki-Haddouti, Christiane: Kryptodebatten. Der Kampf um die Informationshoheit; in: Christiane Schulzki-Haddouti (Hg.): Bürgerrechte im Netz, Bundeszentrale für politische Bildung, Bonn, 2003, S. 162-177

[Ruh05] Ruhmann, Ingo: Cyber-Terrorismus. Panikmache oder reale Gefahr? In: Ulrike Kronfeld-Goharani (Hg.): Friedensbedrohung Terrorismus. Ursachen, Folgen und Gegenstrategien. Kieler Schriften zur Friedenswissenschaft, Band 13, Kiel, 2005, S. 222-240, S. 226

[Ruh14] Ruhmann, Ingo: NSA, IT-Sicherheit und die Folgen. Eine Schadensanalyse; in: DuD Nr. 1, 2014, S. 40-46

[San12] Sanger, David E.: Obama Order Sped Up Wave of Cyberattacks Against Iran; New York Times, June 1, 2012, p. A1; http://www.nytimes.com /2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html

[Sch13] Schmitt, Michael N. (ed.): The Tallinn Manual on the International Law Applicable to Cyber Warfare. Cambridge, 2013

[Sen86] U.S. Senate Select Committee on Intelligence: Meeting the Espionage Challenge: A Review of United States Counterintelligence and Security Programs, Washington, 1986, http://www.intelligence.senate.gov/pdfs99th/99522.pdf

[Spi13] Schnüffelsoftware XKeyscore: Deutsche Geheimdienste setzen US-Spähprogramm ein, in: Spiegel Online, 20.07.2013, http://www.spiegel.de/politik/deutschland/bnd-und-bfv-setzen-nsa-spaehprogramm-xkeyscore-ein-a-912196.html

[UNID13] United Nations Institute for Disarmament Research (UNIDIR) The Cyber Index. International Security Trends and Realities, S. 3; http://www.unidir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Möglichkeiten und Grenzen zur Bestimmung von Cyberwaffen

Thomas Reinhold[1]

Abstract: Seit der Entdeckung der Schadsoftware Stuxnet ist der Cyberspace in den Fokus der internationalen Sicherheitspolitik gerückt. Während Staaten zunehmend die neue Domäne in ihre Sicherheits- und Militärdoktrinen aufnehmen, verdeutlichen Vorkommnisse wie der Sony-Hack oder die Beschädigung eines deutschen Stahlwerks die komplexen Abhängigkeiten von IT-Systemen und deren Verwundbarkeiten. Internationale Bemühungen um die Etablierung verbindlicher Regelungen für das staatliche und militärische Agieren im Cyberspace werden durch ein fehlendes gemeinschaftlich akzeptiertes Verständnis des Themas oder der Definition von Begrifflichkeiten erschwert. Insbesondere das Konzept der "Cyberwaffe" und eine tragfähige und im Kontext international verpflichtender Konventionen belastbare Eingrenzung dieses Begriffs ist dabei für Abrüstungs- und Rüstungskontrollabkommen von zentraler Bedeutung. Eine Gegenüberstellung unterschiedlicher, sowohl generalisierender als auch situationsbezogener Definitionsansätze von Cyberwaffen verdeutlicht dabei die verschiedenen zu integrierenden Betrachtungsebenen. Die im Gegensatz zu konventionellen Waffentechnologien spezifischen Eigenschaften von Software zeigen aber auch die Schwierigkeiten und Grenzen derartiger Ansätze auf.

Keywords: Cyberpeace, Cyberwar, Cyberwaffe, Abrüstung, Rüstungskontrolle

Einleitung

Mit der Entdeckung der Schadsoftware Stuxnet im Juni 2010 [LANGNER2013] ist der Cyberspace in den Fokus der internationalen Sicherheitspolitik gerückt. Die Diskussionen um die Sabotage der iranischen Atomanreicherungsanlage haben dabei ein wichtiges Licht auf diese, in internationalen Debatten lange Zeit vernachlässigte Domäne geworfen, die zunehmend auch in Sicherheits- und Militärdoktrinen Beachtung findet. Eine Studie des Center for Strategic and International Studies [LEWIS2011] stellte bereits 2011 fest, dass 33 Staaten den Cyberspace als weitere Domäne in ihre militärische Planung integriert haben. Die strategische Ausrichtung umfasst dabei neben den defensiven Aspekten und dem Schutz eigener Infrastrukturen zum Teil auch explizit offensive Maßnahmen wie die Entwicklung von Computer Network Operations (CNO) für den Zugriff auf fremde IT-Systeme oder die gezielte Identifikation und Analyse potentieller Ziele im Cyberspace, wie im Falle der „Presidential policy directive PPD-

[1] Institut für Friedensforschung und Sicherheitspolitik an der Universität Hamburg, Beim Schlump 83
20144 Hamburg, reinhold@ifsh.de

20“ des US-Präsidenten vom Oktober 2012 [USA-GOV2012]. Einer aktualisierten Studie des UN Institute for Disarmament Research [UNIDIR2013] zufolge hat sich die Zahl der Staaten mit militärischen Cyber-Programmen seither auf 40 Staaten erhöht, wovon zehn Staaten eine offensive Cyber-Strategie verfolgen und an der Entwicklung offensiv wirksamer Cybermittel arbeiten. Auf der anderen Seiten haben Stuxnet und weitere Vorfälle der vergangenen Jahre die komplexen und umfassenden Abhängigkeiten von IT-Infrastrukturen und Computersystemen und die unklaren Gefährdungen insbesondere kritischer Infrastrukturen verdeutlicht. Beispiele dafür sind die Hacking-Attacke gegen das Netzwerk von Sony-Pictures-Entertainment von 2014 [FBI2014], gegen den französischen Fernsehsender TV5 [HEISE2015] oder die im BSI-Lagebericht von 2014 beschriebene Sabotage eines deutschen Stahlwerks [BSI2014].

In den internationalen politischen Beziehungen haben diese Aspekte zu Verunsicherungen geführt und an historischen Entwicklungen wie dem kalten Krieg, Rüstungswettläufen in der atomaren, biologischen und chemischen Kampfführung oder die Debatten um die Aufrüstung des Weltalls gerührt. Aus Friedens- und Sicherheitspolitischer Sicht wurden damit viele Fragen zum Cyberspace aufgeworfen die bis heute kaum geklärt sind. Obgleich entscheidende internationale Organisationen wie die UN oder die OECD Expertengruppen einberufen und Gremien gegründet haben um sich diesem Problemen zu widmen, werden internationale Debatten oft durch ein fehlendes gemeinschaftlich akzeptiertes Verständnis des Themas oder der Definition von Begrifflichkeiten erschwert. Dazu zählt die Frage nach einer genauen Abgrenzung des Raumes “Cyberspace” und den Aspekten der nationalstaatlichen Souveränität in dieser Domäne ebenso, wie die Definition von Cyberattacken in Verbindung mit der Frage nach der exakten Ausprägung von Cyberangriffen, die als Attacken im Sinne des Völkerrechts verstanden werden können. Exemplarisch dafür steht ein gemeinsamer erster Entwurf der russischen und chinesischen Regierung für eine Cyberkonvention [RU2011]. Darin ist primär die Sprache von “Informationssicherheit” als wesentliches Ziel einer solchen Konvention, die neben der Unverletzlichkeit nationaler IT-Infrastrukturen auch die nationale Souveränität und Kontrolle über die in nationalen IT-Systemen übertragenen und gespeicherten Informationen umfasst. Dieser Ansatz kollidiert dabei mit den Auffassungen von Staaten, die demokratische Werte wie das Recht auf Meinungsfreiheit durch solche Kontrollmöglichkeiten bedroht sehen. Weitere Versuche sich der der gemeinsamen Definitionsfindung zu widmen, wie das Tallinn-Manual [NATO2013] des NATO-Exzellenzzentrums CCDCOE, haben dabei deutlich gemacht, dass etablierte Konzepte des Völkerrechts oder Analogien zu den Diskussionen um klassische militärische Mittel aufgrund spezifischer Eigenschaften des Cyberspace und von Software nur eingeschränkt möglich sind und an ihre Grenzen stoßen.

Normen und Definitionen für den Cyberspace

Die Bemühungen um klare Definitionen stellen daher eines der wichtigsten Grundlagen für die weitere friedliche Ausgestaltung des Cyberspace dar. Sie bilden die Basis für die Entwicklung verbindlicher Normen für staatliches und militärisches Agieren im

Cyberspace und die Festschreibung derartiger Regelungen in Form internationaler Konventionen nach dem Vorbild des Übereinkommen über das Verbot biologischer Waffen [UN1972] oder dem Übereinkommen über das Verbot chemischer Waffen [UN1993]. Derartige Konventionen können in der Vereinbarung von Rüstungsbegrenzenden Maßnahmen (Abrüstung) oder in Übereinkommen über die Kontrolle der Herstellung, des Handels und der Verbreitung von Waffen oder kritischen Technologien und Waffenfähigen Materialien bestehen (Rüstungs- und Exportkontrolle). Insbesondere Abrüstungsabkommen werden in aller Regel gestützt durch die Vereinbarung von Verifikationsregimen, deren Aufgabe unter anderem in der gegenseitigen Überwachung bei der Einhaltung der vereinbarten Regelungen besteht. Solche Verifikationsregime, wie beispielsweise der Atomwaffensperrvertrag (Treaty on the Non-Proliferation of Nuclear Weapons [NPT1968]), beruhen dabei oft auf der Festsetzung von Höchstgrenzen für spezifische Schlüsseltechnologien oder Waffenbestandteile, die durch Inspektionen überwacht werden können. Der Versuch einen solchen Ansatz auf den Cyberspace zu übertragen gestaltet sich besonders aufgrund der schwierigen Definition des Begriffs "Cyberwaffe" und den, im Gegensatz zu konventionellen Waffen spezifischen Eigenschaften von Software als immaterielle, virtuelle und beliebig duplizierbare Produkte problematisch. Als "Cyberwaffe" wird in aller Regel Software jeglicher Art verstanden, die aufgrund ihrer spezifischen Konstruktion darauf angelegt ist, in einem IT-System oder in angeschlossenen Systemen reguläre Abläufe zu stören oder diese über Modifikationen zu zerstören. Diese Betrachtungsweise erfasst jedoch wesentliche Aspekte einer völkerrechtlich hinreichenden Definition einer Waffe, wie den möglichen Zerstörungsumfang oder die Intention eines Angreifers nicht. Zum anderen schließt sie große Bereiche von Software ein, die im zivilen und eindeutig nicht-militärischen Bereich Anwendung finden. Ein engere Umgrenzung des Begriffs ist daher geboten.

Definitionsversuche zu Cyberwaffen

Die Autoren der OECD-Studie "Reducing Systemic Cybersecurity Risk" [SOMMER2010] widmen sich der Frage nach der Definition von Cyberwaffen unter dem Blickwinkel der Eigenschaften klassischer Waffen: "*There is an important distinction between something that causes unpleasant or even deadly effects and a weapon. A weapon is "directed force" – its release can be controlled, there is a reasonable forecast of the effects it will have, and it will not damage the user, his friends or innocent third parties"*. Um insbesondere die bei Cyberwaffen problematische zeitnahe Zuordnung eines Angriffs - das sogenannte Attributionsproblem - aufzugreifen schlagen die Autoren bei der Bewertung von Schadsoftware weitere Bewertungsaspekte vor und empfehlen eine Analyse anhand der folgenden Kriterien:

> *Is this something whose targeting and impact can be controlled (is there a risk of friendly fire?)*
> *What success rate can be expected in terms of targets?*
> *Is there any collateral damage?*

What resources and skills are required?
How much inside knowledge and/or inside access of target is required? How easy is this to achieve?
Can the weapon be detected before or during deployment?
Can a perpetrator be detected during or after deployment?
What are the actual effects and how long do they last?
How long can an attack be carried out before it is thwarted by counter-technology?
How long can an attack be carried out before perpetrators are identified?

Die Autoren definieren damit ein wirksames und auf Software übertragbares Raster anhand dessen sich Schadsoftware auf Basis der spezifischen technischen Umsetzung voneinander abgrenzen lässt. Allerdings nehmen die Autoren bei ihrer Aufstellung der Bewertungsmaßstäbe keine Gewichtungen der einzelnen Kriterien vor und vermeiden eine klare Aussage über die Unterscheidung zwischen krimineller Schadsoftware und Cyberwaffen. Die Kriterien sollen in erster Linie als Hilfsmittel dienen um konkrete Vorfälle zu bewerten. Daher eignen sie sich insbesondere um die Fülle an verfügbarer Schadsoftware bei der Diskussion um Cyberwaffen einzugrenzen: "*On this basis it will be seen that the most common forms of virus (..) fail as credible cyberweapons, because they are relatively difficult to control. However a targeted DDoS is a likely cyberweapon.*"

Ein weiterer Versuch der Begriffsfindung besteht in der Betrachtung des tatsächlich bewirkten Schadens einer Schadsoftware. Dieser kann aufgrund der starken Vernetzung sowie der komplexen und verzögerten Wechselwirkungen von IT-Systemen unter Umständen beträchtlich von der intendierten Wirkung abweichen, wie unzählige Vorfälle der vergangenen Jahre gezeigt haben. So wurde die Schadsoftware Stuxnet beispielsweise auf sehr viel mehr Rechner weltweit entdeckt als nur auf den IT-Systemen der Anreicherungsanlage in Natanz, die Stuxnet sabotieren sollte. Den Ansatz Schadsoftware über die tatsächlich ausgelöste Wirkung zu kategorisieren verfolgen die Autoren der Studie "On the Spectrum of Cyberspace Operations" [BROWN2012]. Sie bewerten den beabsichtigten und unbeabsichtigten Schaden gemeinsam innerhalb eines Spektrums, das von dem reinen Eindringen und nicht-invasiven Agierens in fremden Computersystemen ("*enabling operations*") über die zeitweise Unterbrechung von Diensten ("*cyber disruption*") bis hin zur tatsächlichen Schädigung von Dingen oder Personen reicht ("*cyber attacks*"). Aus Sicht der Autoren entsprechen dabei erst Varianten von Schadsoftware, deren Auswirkung in letztere Kategorie fallen, dem Konzept einer Waffe beziehungsweise eines bewaffneten Angriff im Sinne des Völkerrechts nach der UN Charta Art. 51 [UN1945]. Auch dieser explizit situationsbezogene Ansatz eignet sich damit eher für die Bewertung konkreter Vorfälle, insbesondere für die Abwägung nach angemessenen Reaktionen eines angegriffenen Staates, wie für die im Recht auf Selbstverteidigung vorgesehene Maßgabe der Proportionalität von Maßnahmen. Obgleich für eine Definition im Rahmen internationalen Konventionen eher die Betrachtung der strukturell intendierten Wirkung einer Schadsoftware nötig ist, so bietet sich die Kontinuums-Klassifikation doch an um

große Teil der "üblichen" Hacking-Attacken wie dem Einbruch in IT-Systeme und dem Diebstahl von Daten für die Begriffsfindung auszuschließen und in den Bereich der nationalen und internationalen Strafverfolgung zu verweisen.

Eine dritte Studie "Cyber-weapons: legal and strategic aspects" [MELE2013] betrachtet die Definition einer Cyberwaffe unter rein konzeptionellen Aspekten: "*a weapon can be also an abstract concept thereby not necessarily a material one, as international and domestic legislation have considered it up to now*". Damit berücksichtigt der Autor den Umstand, dass sich Schadsoftware, die für kriminelle Zwecke benutzt wird, von einem technischen Standpunkt aus betrachtet kaum von Schadsoftware unterscheiden muss, deren Ziel die Zerstörung eines IT-Systems ist. Beide Varianten benötigen Mechanismen um in ihr Zielsystem einzudringen, sich vor Abwehrmechanismen zu schützen und ihren eigentlichen Schadmechanismus auszuführen. Ob dieser im Entwenden von Daten oder dem vollständigen Löschen eines IT-Systems besteht liegt dabei primär in der Intention des Angreifers und weniger in technischen Spezifika der Schadsoftware. Der Autor schlägt für die Eingrenzung einer Schadsoftware als Cyberwaffe daher die Berücksichtigung juristischer und strategischer Dimensionen vor. Diese umfassen den Anwendungskontext und den Zweck eines Schadsoftware-Einsatzes, sowie den beabsichtigten Schaden und die konkrete absichtsvolle Auswahl eines strategisch relevanten Ziels: "*[a cyberweapon is] a part of equipment, a device or any set of computer instructions used in a conflict among actors, both National and non-National, with the purpose of causing, even indirectly, a physical damage to equipment or people, or rather of sabotaging or damaging in a direct way the information systems of a sensitive target of the attacked subject (..) with the purpose of achieving, keeping or defending a condition of strategic, operative and/or tactical advantage.*" Ein solcher Ansatz, der vor allem die Bewertung eines Vorfalls, des vermuteten Angreifers und dessen Intentionen anstelle des tatsächlichen Geschehens und reeller Schäden in den Fokus rückt, entspricht damit in hohem Maße dem allgemeinen Umgang mit Cybervorfällen durch staatliche Institutionen. Dies wird exemplarisch deutlich am Beispiel der Hacking-Attacke gegen das Netzwerks des in den USA ansässigen Unternehmens Sony-Picture-Entertainment [FBI2014]. Trotz zweifelhafter offizieller Beweise für die Herkunft der Angreifer - mutmaßlich militärische Einheiten aus Nord-Korea - und geringer nachgewiesener Schäden - Diebstahl von Daten - wurde der Vorfall durch die US-Regierung als Bedrohung ihrer inneren Sicherheit gewertet und ausgewählte nordkoreanische Unternehmen und Personen umgehend mit wirtschaftlichen Sanktion belegt [USTREASURY2015].

Eines der hervorgehobenen Bewertungskriterien des konzeptionellen Ansatzes betrifft die gezielte Entwicklung von Cyberwaffen für konkret ausgewählte strategisch relevante Ziele. Mit Blick auf den Aufwand eines solchen Vorhabens kommt der Autor hinsichtlich der weltweiten Verfügbarkeit von Cyberwaffen zu folgender Schlußfolgerung: "*Summarizing these observations, it is possible to understand how the creation and the employment of cyber-weapons require superior intelligence information, time, workforce and testing resources for their creation (..) and ensuring the possibility*

to hit targets often unreachable by other types of attacks (..) The high costs, the risk variables for their creation and efficiency, as well as the "limited" and anyway temporary results, lead to believe that currently research and development activities in the field of cyber-weapons are strategically unprofitable, unless an escalation takes place (..) in the power these software have to increase the damaging level and/or to make their effects last as long as possible".

Die Gegenüberstellung der unterschiedlichen Herangehensweisen verdeutlicht die unterschiedlichen Betrachtungsebenen als Klassifikationsmöglichkeiten von Cyberwaffen, von der Bewertung der technischen Eigenschaften einer Software, über den tatsächlichen Schaden bis zum konzeptionellen und strategisch geplanten Zweck der Software. Zusammengefasst ergeben diese Sichtweisen die folgenden Fragen:

Wie ist die Software exakt technisch gestaltet
Wofür ist die Software strategisch gedacht
Wogegen wurde die Software eingesetzt
Was ist der tatsächlich verursachte Schaden am Ziel und darüber hinaus

Die drei vorgestellten Ansätze verdeutlichen, dass Cyberwaffen trotz spezifischer Eigenschaften, die sie von bisherigen Waffen unterscheiden und auf die im folgenden noch näher eingegangen wird, mit Hilfe etablierter Konzepte für die Kategorisierung klassischer Waffen betrachtet werden können. Damit können diese eine wichtige Grundlage für die Entwicklung von Cyberkonventionen und Abkommen bilden. Dies gilt insbesondere dann, wenn weniger die konkrete technische Ausprägung als vielmehr der beabsichtigte Schaden und der strategische Anwendungszweck im Mittelpunkt stehen. Darüber hinaus zeigen die Ansätze, welche Formen von Schadsoftware vermutlich nicht als Waffe im Sinne des Völkerrechts gelten können, selbst wenn sie im Einzelfall Schäden anrichten. Sie dienen damit der wichtigen Unterscheidung zwischen kriminellem Handeln, dem mit Mitteln der Strafverfolgung begegnet werden kann, und staatlich militärischen Aktivitäten, die nur auf der Ebene von verbindlichen bilateralen oder internationalen Vereinbarungen geregelt werden können.

Das Wassenaar-Abkommen als erster Schritt einer Cyberwaffen-Konvention

Ein erster Ansatz Cyberwaffen im Rahmen einer internationalen Vereinbahrung zu erfassen und deren Verbreitung zu regulieren wurde durch eine Ergänzung des "Wassenaar-Abkommens für Exportkontrollen von konventionellen Waffen und doppelverwendungsfähigen Güter und Technologien" [WASSENAAR1995] unternommen. Dieses, seit 1996 bestehende Abkommen dient der Aufgabe innerhalb der Gruppe von Mitgliedsstaaten Transparenz über den Export mit kritischen und waffenfähigen Ressourcen und Schlüsseltechnologien zu schaffen. Die Wassenaar-Vereinbarungen enthalten jedoch keine Übereinkünfte zu Rüstungsbegrenzenden Maßnahmen und tragen damit zwar den Charakter der Exportkontrolle, nicht jedoch der

Abrüstung. Ende 2013 wurde die Liste der zu kontrollierenden Technologie um "*intrusion software*" erweitert [WASSENAAR2013], mit folgender Definition:

"Software" specially designed or modified to avoid detection by 'monitoring tools', or to defeat 'protective countermeasures', of a computer or network capable device, and performing any of the following:

The extraction of data or information, from a computer or network capable device, or the modification of system or user data; or

The modification of the standard execution path of a program or process in order to allow the execution of externally provided instructions.

Mit Blick auf die drei vorgestellten Klassifikationsansätze fällt auf, dass in der Wassenaar-Definition weder eine nähere Einschränkung auf einen strukturell intendierten Schaden der Software noch deren strategischer Einsatzzweck vorgenommen wird. Ausschließlich die technologisch bereitgestellten Fähigkeiten zum Stören oder Zerstören sowie zur Extraktion von Daten aus IT-Systemen gelten als hinreichende Abgrenzung. Diese Definition ist sehr weit gefasst und überläßt den Mitgliedsstaaten, in deren Verantwortung es liegt Exportbestimmung individuell in nationalem Recht zu verankern, einen breiten Interpretations- und Handlungsspielraum. Für eine internationale Konvention, die über den Kreis der gegenwärtig 41 Mitgliedsstaaten des Wassenaar-Abkommens hinaus, geht wird sich eine solch allgemeine Definition nicht durchsetzen lassen, da sie keine einheitliche Bewertungsgrundlage schafft und so die Vergleichbarkeit von nationalen Verfahrensstandards erschwert.

Erschwerende Spezifika von Schadsoftware

Neben den erläuterten Schwierigkeiten bei der Abgrenzung von Cyberwaffen existieren weitere Rahmenbedingungen die im Vergleich zu konventionellen waffenfähigen Technologien für Schadsoftware einzigartig sind und bei der Definition von Cyberwaffen berücksichtigt werden müssen. Dies betrifft zum einen die fehlende Materialität und die beliebige Dublizierbarkeit von Software, die jegliche effektive Kontrolle von Schadsoftware anhand der Begrenzung des Umfangs von Waffen-Arsenalen oder deren geographischer Lokalisierung verhindert. Diese Maßnahmen sind für klassische Nichtverbreitungs-Verträge sowie für die Nachverfolgbarkeit des Handels mit kritischen Technologien maßgeblich und wichtige Grundlage für Abrüstungs- und Friedensinitiativen. Entsprechende Vereinbarungen wie demilitarisierte Zonen oder die Beschränkung auf Maximalmengen bestimmter Cyberwaffen scheiden daher aus. Ein weiteres Problem bei der Klassifikation einer konkreten Software besteht in der Abgrenzung ihres Zwecks zwischen ausschließlich defensiven und potentiell offensiven Fähigkeiten. Während defensiv orientierte Maßnahmen völkerrechtlich eher unstrittig sind, ist die Abgrenzung defensiv-offensiv für Software schwer vorzunehmen. IT-Systeme effektiv auf Schwächen zu testen umfasst beispielsweise die Entwicklung und den Einsatz von Software für sogenannte Penetrations-Tests, bei denen Angriffe auf eigene Systeme gezielt durchgeführt werden um Schwachstellen zu entdecken, zu

beheben und geeignete Schutzmaßnahmen zu etablieren. Derartige Penetrations-Test können jedoch beispielsweise nicht nur gegen eigene IT-Systeme eingesetzt werden und fallen damit - je nach Interpretation des regulierenden Staates - unter Umständen bereits in die Kategorie der "*intrusion software*" der Wassenaar-Exportkontrollbeschränkungen. Die Regelungen des in Deutschland für die Kontrolle der Wassenaar-Bestimmungen zuständigen Bundesamtes für Wirtschaft und Ausfuhrkontrolle [BAFA2014] folgen einer solchen Interpretation und sehen gegenwärtig die Pflicht für Exportgenehmigungen für einige spezifische Exportländer vor. Das Wassenaar-Abkommen verdeutlicht indes ein weiteres Abgrenzungsproblem bei Schadsoftware. Es wurde insbesondere für die Kontrolle von kritischen Technologien und Gütern geschaffen, die sowohl zivilen als auch militärischen Zwecken dienen können, um aufgrund dieses sogenannten Dual-Use-Charakters im Einzelfall über Exportbeschränkungen entscheiden zu können. Eine solche Entscheidung ist für Software ist jedoch nur schwer vorzunehmen, da wie dargelegt maßgeblich die Intention des Anwenders über die Wirkung des Einsatzes einer Software entscheidet und nicht primär deren technisches Potential. Ein weiteres Problem besteht in der Kontrolle der Verbreitung von Sicherheitslücken in IT-Systemen, die meist den Grundstein einer Schadsoftware bilden und mit Hilfe dessen eine Software Sicherheitsmechanismen umgehen und sich Zutritt zu geschützten IT-Systemen verschaffen kann. Derartige Sicherheitslücken für populäre Software haben teilweise einen hohen ökonomischen Wert und es ist oft lukrativer mit entdeckten Schwächen zu handeln als diese an den Hersteller der Software für eine Fehlerbehebung zu melden. Darüber hinaus bedeutet die Entdeckung schwerwiegender Sicherheitslücken in populärer Software oft einen Image-Schaden für den Herstellers weshalb auch seitens der Unternehmen Schwächen in ihren Produkten möglicherweise eher verschwiegen als veröffentlicht und mit Aktualisierungen behoben werden. Aus diesem Grund ist der Markt des "waffenfähigen Materials" für Cyberwaffen schwer zu überschauen oder zu kontrollieren. Eine belastbare und international als zuverlässig anerkannte Definition von Cyberwaffen muss daher eine verbindliche Kontrolle der Verbreitung von Sicherheitslücken, wie es mit dem Wassenaar-Abkommen in einem ersten Schritt unternommen wurde umfassen.

Fazit

Die vorgestellten Klassifikationskonzepte verdeutlichen, dass auch das Konzept von immateriellen Waffen im virtuellen Cyberspace mit Hilfe etablierter Vorgehensweisen der Bewertung von militärischen Technologien erfassbar sind. Die für effektive und konkrete Abkommen notwendigen Regelungen und Definitionsfeinheiten werden aber gegenwärtig zum einen noch durch unterschiedliche Herangehensweisen und zum anderen durch die für Software einzigartigen Eigenschaften behindert. Ein tragfähige und im Kontext international verpflichtender Konventionen belastbare und sinnvolle Eingrenzung des Begriffs Cyberwaffe muss daher zum einen diese verschiedenen Betrachtungswinkel integrieren. Zum anderen verdeutlichen der ausgeprägte Dual-Use-Charakter und die ambivalente Abgrenzung von defensiven und offensiven Fähigkeiten einer Software, dass es für Cyberwaffen mehr als für bisherige klassische Waffen-

Technologien auf die Einzelfallprüfung ankommen wird. Die erweiterten Regelungen des Wassenaar-Abkommens können hier einen ersten sinnvollen Versuch darstellen über den sich Staaten zu weiteren und weitreichenderen Schritten auf dem Weg einer Kontrolle und Eingrenzung der militärischen Nutzung des Cyberspace austauschen und verpflichten könnten.

Literaturverzeichnis

[BAFA2014] Bundesamt für Wirtschaft und Ausfuhrkontrolle, "Ausfuhrliste B, National erfasste Güter", 2014

[BROWN2012] Brown, G. D. & Tullos, O. W., "On the Spectrum of Cyberspace Operations", Small Wars Journal, 2012

[BSI2014] Bundesamt für Sicherheit in der Informationstechnik, "Die Lage der IT-Sicherheit in Deutschland 2014", Bonn, 2014

[FBI2014] Federal Bureau of Investigation, "Update on Sony Investigation", Washington, 2014

[HEISE2015] Heise.de, "Islamisten hacken TV5", http://heise.de/-2597578, 2015

[LANGNER2013] Langner, R. „To Kill a Centrifuge - A Technical Analysis of What Stuxnet's Creators Tried to Achieve", Hamburg, 2013

[LEWIS2011] Lewis, J. A. & Timlin, K., Cybersecurity and Cyberwarfare – Preliminary Assessment of National Doctrine and Organization, Center for Strategic and International Studies, 2011

[MELE2013] Mele, S., "Cyber-weapons: legal and strategic aspects", Italian Institute of Strategic Studies "Niccolò Machiavelli", 2013

[NATO2013] NATO Cooperative Cyber Defence Centre of Excellence (NATO CCDCOE), „The Tallinn Manual on the International Law Applicable to Cyber Warfare", Tallinn, 2013

[NPT1968] Treaty on the Non-Proliferation of Nuclear Weapons, 1968

[RU2011] Ministry of Foreign Affairs of Russia "Convention on international information security", 2011

[SOMMER2010] Sommer, P. und Brown, I., „OECD Study - Reducing Systemic Cybersecurity Risk OECD/IFP Project on Future Global Shocks", 2010

[UN1945] Vereinte Nationen, „Charta der Vereinten Nationen und Statut des Internationalen Gerichtshofs", 1945

[UN1972] Vereinte Nationen, "Übereinkommen über das Verbot der Entwicklung, Herstellung und Lagerung bakteriologischer (biologischer) Waffen und von Toxinwaffen sowie über die Vernichtung solcher Waffen", 1972

[UN1993] Vereinte Nationen, "Übereinkommen über das Verbot der Entwicklung, Herstellung, Lagerung und des Einsatzes chemischer Waffen und über die Vernichtung solcher Waffen", 1993

[UNIDIR2013] UNIDIR, United Nations Institute for Disarmament Research, „The Cyber Index - International Security Trends and Realities", Genf, 2013

[USA-GOV2012] US-Regierung, Barack Obama, „Presidential policy directive / PPD-

20“, Washington, 2012
[USTREASURY2015] US Department of treasury, “Treasury Imposes Sanctions Against the Government of The Democratic People’s Republic Of Korea”, 2015
[WASSENAAR1995] “The Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies”, Wassenaar, 1995
[WASSENAAR2013] “The Wassenaar Arrangement on export controls for conventional arms and dual-use goods and technologies - List of dual-use goods and technologies and munitions list”, 2013

Strategic Operations in the Cyber Domain and their Implications for National Cyber Security[1]

Sven Herpig[2]

Abstract: With the public discussion mainly revolving around deterrence (either by denial or by punishment), there are actually a number of strategies which can be applied to cyber operations. A cyber strategy can be thought of as an umbrella for various individual cyber operations with the ultimate aim to achieve a strategic and / or political goal. Thus, cyber strategies can be defined as the development and employment of cyber operations, integrated and coordinated with other operational domains and forms of information warfare, to achieve or support the achievement of political objectives. Five of those strategies currently exist: going dark, deterrence, sub rosa, shashou jian, and cyber war. The implications of their existence and use create the need for proactive cyber security.

Keywords: Cyber Security, Cyber Operations, Cyber Strategies, Cyber Warfare, Deterrence, Cyber War

1 Introduction

Over the last centuries, the world saw economically interwoven countries and a tendency to lower the use of force moving towards veiled warfare. It is one of the few tools which can be used to conduct actions against an adversary without necessarily sparking chaos in the international arena. Indeed, it seems that cyber operations are currently en vogue. The cyber campaign Olympic Games, commonly referred to only as Stuxnet and its relatives, is the poster-boy of the development. Various cyber weapons were developed to work together in order to infiltrate, analyse, sabotage and ultimately erase their traces [BPBF12][Gb12][FMC10]. Olympic Games hit Iran's centre of nuclear activity, a research and production facility which otherwise could have only been affected by a physical attack (e.g. a precise air strike). The latter would have caused the death of many people and might have destabilized this region even further.

Even though Robinson, Jones and Janicke [Rm15] researched and listed a comprehensive list of current research challenges for cyber warfare, they did not include cyber strategies as one of them. With the public discussion mainly revolving around deterrence (either by denial or by punishment), there are actually a number of strategies which can be applied to cyber operations. Knowing the various options is vital because

[1] Edited and abbreviated excerpt based on the PhD thesis 'Anti-War and the Cyber Triangle: Strategic Implications of Cyber Operations and Security for the State' submitted to the University of Hull by Sven Herpig in 2014. Currently pending for final approval.

[2] Bundesamt für Sicherheit in der Informationstechnik, Godesberger Alle 185-189, 53175, Bonn, sven.herpig@bsi.bund.de

'[a] grand strategic vision of cyberspace can assist states in navigating the informational turbulence in which contemporary international politics appears to find itself. [...] Cyberspace has its myriad of problems, but a true strategic sensibility demands that long-term interests prevail over short-term opportunism' [BS11]. A cyber strategy can be thought of as an umbrella for various individual cyber operations with the ultimate aim to achieve a strategic and / or political goal. Kuehl identifies a cyber strategy as 'the development and employment of strategic capabilities to operate in cyberspace, integrated and coordinated with the other operational domains, to achieve or support the achievement of objectives across the elements of national power in support of national security strategy' [Kd09]. Starr sees it as 'the development and employment of capabilities to operate in cyberspace, integrated and coordinated with the other operational domains, to achieve or support the achievement of objectives across the elements of national power' [Ss06]. Thus, cyber strategies can be defined as 'the development and employment of cyber operations, integrated and coordinated with other operational domains and forms of information warfare, to achieve or support the achievement of political objectives. Cyber operations refer to the targeted use and hack of digital code by any individual, group, organization or state using digital networks and connected devices, which is directed against critical national, military or civilian information infrastructure in order to alter, destroy, disrupt or deny functionality with the ultimate aim to weaken and/ or harm the targeted political unit' [Hs15].

Thus, a comprehensive cyber strategy consists inter alia of a certain implementation of cyber security as well as all the cyber operations carried out under its umbrella, connecting them to achieve a particular political or strategic goal. Discussions about the general idea of cyber strategies started with Libicki's works on deterrence and have been developed ever since [Lm09a][Lm09b] but also go far beyond that. This paper identifies and discusses the five existing cyber strategies on a macro level and analyses their implications with focus on national cyber security. The discussion is based on the academic examination of 50 cyber operations from the perspective of strategic studies and a case study of the *Olympic Games* campaign (empirical contribution) as well as extensive research on the subject matter (theoretical contribution) including more than 300 sources on strategy, cyber warfare and related areas [Hs15]. It shall therefore provide a strategic umbrella for bridging the gap between the more technical nature of cyber operations and policy-maker understanding of their implications.

2 Cyber Strategies

2.1 Going Dark

This strategy is an extended and broader implementation of the security mechanism of air-gaping networks. Going Dark means that all systems and networks which are part of the (critical) national information infrastructure of a country are not connected neither to wider networks nor to the Internet. Gervais mentions this strategy, stating that: '[...] when

it comes to states, like North Korea, that are less technologically advanced, cyber reprisals have little effect. Reprisals to cyber attacks, therefore, ought to manifest themselves as physical countermeasures when necessary' [Gm11]. It can also mean that there is no such an infrastructure to speak of yet, due to the low level of development of this state. Relying on this strategy shows that the adopting state does not believe in its ability to defend its networks properly – or their vitality, therefore making them go dark completely. Implementing this strategy can be done partially – only the classified networks are air-gaped – or for the active networks and systems structure. This approach tries to deny the adversary any access to the systems and networks which hold valuable data or can lead to casualties. The delineation between going dark and deterrence by denial is the fact that going dark does not include any additional measures of protecting systems and networks, whereas deterrence by denial includes a comprehensive set of hardening activities (see 2.2). The Olympic Games campaign has been an implementation of a shashou jian strategy (compare 2.4) consisting of different operations which targeted an infrastructure which can be considered as 'gone dark' – and partially succeeded [FMC10].

2.2 Deterrence

Deterrence in the cyber domain is the most developed and analysed cyber strategy today, if not the only one that has been discussed thoroughly so far. In general, deterrence is subdivided into deterrence by denial and deterrence by punishment. Deterrence by denial is '[...] to deny an adversary the ability to achieve its military and political objectives [...]' [Gs61] whereas '[t]he goal of deterrence by punishment is to prevent aggression by threatening greater aggression in the form of painful and perhaps fatal retaliation' [Gk11].

The most important point of deterrence is its credibility [Kr09]. The adversary has to believe that the opponent's threat of retaliation is credible. If the adversary believes it, he will not attack and is therefore deterred, or as Gray put it: '[t]he deterree has to agree to be deterred, no matter how unwillingly' [Gc99]. One common misconception about cyber deterrence is to be highlighted first. Cyber operations are used as a means to deter any domain adversarial aggression. Cyber deterrence does not mean the use of any domain means to deter an adversarial cyber aggression. For example: the threat of use of nuclear weapons as retaliation for a cyber attack is not cyber deterrence but nuclear deterrence. If both stakeholders then implement cyber retaliation, it might lead to a 'mutually assured disruption' [Gk11].

Thus, deterrence in the cyber domain may need offensive and defensive capabilities to be in place at the same time to create credible deterrence. However, airtight defensive security could make up for the lack of offensive capabilities. The ultimate aim subsequently is to increase the own security. Kugler states that '[...] the potential payoff of a well-conceived cyber deterrence strategy is considerably greater security than exists today' [Kr09]. Deterrence using cyber operations can be implemented in various ways. Payne and Walton define three types of deterrence 1. deterrence as direct attack, 2.

deterrence as preventing from doing a provocative act and 3. aggression becomes unprofitable [PW02]. Kugler's suggestions are similar. He states that the three types of how deterrence in the framework of cyber operations could work are 1. deterrence by denying benefits, 2. deterrence by incentives as well as 3. deterrence by imposing costs [Kr09].

Deterrence is not an either-or decision. Strategies can all work at once, or equally fail together. Therefore, Starr suggested a concept where cyber deterrence is custom-tailored to the adversary [Ss09]. In the case of states, Starr suggests to carry out cyber espionage activities against them to be able to tailor the deterrence strategy. This conclusion can be derived from the nature of cyber armoury. If state A were able to penetrate the networks of state B, it can be assumed that malicious software has been planted, the perception is a persuasive here as actuality. Therefore, state B might be deterred from attacking A because it assumes that A can detonate those time bombs any time. In addition, A might have gained knowledge about the weapons that B has and can harden and shield its networks from likely retaliation which effectively render B's potential attacks useless. A might have even found more vulnerabilities to exploit B's networks for future endeavours. Therefore, credible cyber deterrence needs to be custom-tailored and relies on information acquired through intelligence operations.

Despite the opportunities mentioned, cyber deterrence faces several challenges and some authors therefore regard it as void [CK10][LX10]. Lewis for example states that, '[t]he fundamental assumption is that a correct interpretation by opponents will lead them to reject certain courses of action as too risky or too expensive. The problem is that potential opponents may misinterpret deterrent threats while others may be not feel threatened, and are therefore harder to deter' [Lj10]. In case of cyber deterrence against cyber attacks, the primary challenge is proper attribution, mentioned earlier this chapter, which undermines the credibility of cyber deterrence to a large degree. If an attacker cannot be properly identified, it cannot produce deterrence by punishment. Therefore cyber deterrence might fail. If an attacker can be identified it can still be an act of deception. There can be no 100% proof whether the clues leading to the possible attacker are genuine or as distraction as part of a deceiving cyber operation. Therefore, cyber deterrence would also fail as long as no perpetrator officially takes responsibility for the attack. Even then, terrorists for example might claim ownership of a cyber attack to spread terror, whereas the actual attacker does not want to make his involvement public. Hence, the lack of proper attribution is a large problem for credibility, and hence successful deterrence. Thus, in case of an attack and subsequent possible retaliation, the decision is deferred to the political level.

Another challenge is that some cyber attacks might be too small to retaliate against [Kr09] and subsequently undermine a zero tolerance policy [He10]. If a state communicates that it will retaliate against every cyber attack (zero tolerance) it is doomed to fail because of the sheer number of attacks and the lack of resources to respond to them. Having declared retaliation against every cyber attack but failing to do so, undermines a state’s credibility. The political level therefore has to set and communicate a threshold: how much damage a cyber attack has to do for a cyber

retaliation to trigger and therefore deterrence to take place. All cyber attacks below this threshold do not produce the effect of deterrence. The other option would be the implementation of a zero tolerance policy which would inadvertently fail and therefore diminish the credibility of cyber deterrence.

Cyber deterrence against cyber attacks (as well as other attacks) struggles to deal effectively with non-state stakeholders [Lm09b]. While not covered by this research, this potential challenge warrants mention. The threat of wiping an individual's computer is not credible enough to prevent him from trying to shut down the power grid of a country. This leads to the next challenge, the lack of impact in case of cyber retaliation against a state. Compared to nuclear weapons, cyber operations lack the ability for mutual assured destruction [Aj01] or 'unexpected higher- order effects' [Ss09]. If country A plans to invade country B and has a high chance of success, A would unlikely be deterred by B's potential to shut down the power grid and wipe important databases. A is more likely to be deterred, however if B could wipe-out A's capital city as a response to the invasion. Due to the nature of cyber weapons, cyber deterrence as a strategy also faces the problem that most cyber weapons are one use only [He10]. They exploit vulnerabilities and once the adversary notices, he can fix the vulnerability and therefore render the weapon useless (against him). The knowledge of this increases the threshold of retaliation for the deterrer owing to hesitancy to use up his cyber arsenal. This increases the threshold to a level that retaliation as a result from cyber deterrence always borders between escalation and impunity, [Kr09] or as Hjortdal puts it '[t]he strategy of deterrence is thus two-sided and, as such, contradictory—a balancing act is needed between hiding the maximum level of capability on the one hand, and communicating and proving that the capability exists on a sufficiently high level to deter other states on the other' [Hm11], a thin line. For further research on this issue, when taking into account a multi- stakeholder setting, cyber deterrence faces the challenge of extended cyber defense and collective cyber retaliation only to work if applied sub rosa but not publicly [Lm09a].

Sharma sees cyber deterrence as the only vital defence against cyber attacks [Sa09]. This is partly accurate. It is the only viable cyber defence strategy which can be applied across the (critical) national information infrastructure - as opposed to going dark which can only be partly applied. However, cyber deterrence is heavily restricted in what it can achieve.

2.3 Sub Rosa

Extraction and disruption operations using networks and computer system have been coined sub rosa activities by Libicki [Lm09b]. Subsequently, a sub rosa cyber strategy 'has some aspects of intelligence operations, and some aspects of special operations – although it is neither. Of note, sub rosa warfare is almost impossible to conduct with tanks, much less nuclear weapons' [Lm09b]. Sub rosa cyber strategy are covered in some works in a blurred pool of cyber operations, information warfare and intelligence operations, but not often distinctly discussed as a single and genuine strategy or

approach. It bears close resemblance with traditional sub rosa activities such as espionage or sabotage but is conducted through cyber operations. Thus, a sub rosa cyber strategy can be part of a major intelligence operation which also involves other elements such as human intelligence (HUMINT).

States are aware of this strategy, as Gervais suggests when stating that '[a]necdotal evidence suggests that cyber espionage is a familiar practice of state governments' [Gm11]. Betz and Stevens even suggest that sub rosa cyber strategy are aspiring to be the most prevalent cyber strategy, as compared with strategies with a higher level of intensity [BS11].

A sub rosa cyber strategy is only sub rosa as long as both parties agree it to be, or as Libicki phrases it: '[p]aradoxically, maintaining sub rosa warfare requires the tacit assent of the other side, and is therefore quite fragile' [Lm09b]. The reason to keep it secret is that the less the public knows, the easier it is to de-escalate the conflict [Lm09b]. If one of the stakeholders decides to end its secretive conduct, the sub rosa operations, if continued, turn into for example shashou jian strategy (see 2.4). This strategy has a higher level of intensity and therefore does not only mean to turn a covert operation overt, but also to increase the risk of escalation and subsequent retaliation. Keeping operations sub rosa through this strategy means decreasing the likelihood of entering the retaliation cycle [Lm09b]. The more intense and physical sub rosa operations are, the more likely they are to escalate. If state A shuts down state B's power grid, B is politically pressured to react – even more so if the perpetrator becomes public knowledge. The sub rosa cyber strategy is therefore a limited intensity strategy with a likelihood of the involved stakeholders being aware of the operations but deliberately keeping them covert in order to avoid decreasing political leeway.

There is a thin line between a sub rosa cyber strategy and the shashou jian cyber strategy. It is prudent however to differentiate those two strategies from one another for several reasons. Apart from the difference in indicators which are discussed in the respective categorization paragraphs, the core distinction is that the sub rosa strategy mainly refers to intelligence, not sabotage. This crucial element coincides with the covertness of a sub rosa cyber strategy as compared to a potentially overt character of shashou jian operations as acts of sabotage are more difficult to keep covert. Sub rosa, is not, however, anything new. It is covert intelligence operations carried out through the use of cyber operations. Therefore it is necessary to distinguish it from other cyber strategies, it is less necessary to do so from other intelligence operations.

From a state's perspective, it is prudent to start implementing a sub rosa strategy by actually strengthening the own cyber security approach. When engaging in offensive cyber activities, one can expect to be attacked as well – either as retaliation or just because of the assumption that certain attacks can just not be backtracked and subsequently attributed. It is vital to develop the defense at least to the level that it could withstand an attack mirroring the power and effort oneself puts into offensive cyber operations.

2.4 Shashou Jian

Shashou jian is the Chinese translation for assassin's mace, a strategy which refers to the ability of striking the enemy decisively and stealthily - making the fight fit the weapons [CK10][Nl05]. Incorporating this strategy into the cyber operations framework is based on the alleged Chinese use of shashou jian as a means to achieve its geo-strategic goals [Nl05]. The use of the term in this work might exceed the depth of shashou jian in the Chinese original meaning. It seems however useful to keep the term and extend the description as it reflects not only the use by Chinese strategists in general but also the connection of Sun Tzu's teachings to this concept. Sun Tzu describes this kind of strategy in his writings as relying on speed, stating that '[s]peed is the essence of war. Take advantage of the enemy's unpreparedness, travel by unexpected routes and strike him where he has taken no precautions' [SG63]. In conventional terms, an assassin's mace strategy can be pictured as an attacker coming out from cover to deal a swift blow to the victim – and at once disappears.

Libicki discusses three key roles which a cyber attack might play: '[i]t might cripple adversary capabilities quickly, if the adversary is caught by surprise. It can be used as a rapier in limited situations, thereby affording a temporary but potentially decisive military advantage. It can also inhibit the adversary from using its system confidently' [Lm09a]. All the three roles are goals that can be achieved with a shashou jian cyber strategy. It aims at the decisive points [Ja68] or centres of gravity [Rg01] of the enemy to carry out a precise blow, ignoring the rules of conduct [Fj08] to achieve a coup de grâce [Tm67]. One targeted blow against parts of the (critical) national information infrastructure that brings about a huge impact (for example bringing down the state's entire power grid).

Shashou jian is very versatile can be carried out in the framework of warfare or under the umbrella of intelligence operations. When linked to the latter, it is most likely affiliated with sabotage rather than espionage activities. Shashou jian does not necessarily work in supplement to other forms of warfare or intelligence operations, but can be a standalone strategy. Hence, sub rosa and shashou jian are not only cyber strategies, but can also be conducted under the umbrella of intelligence operations. Even if to distinguish between espionage and sabotage activities seems arbitrary, it is not. The genuine difference between sub rosa and shashou jian strategies is that shashou jian still works as an overt operation after it has successfully been carried out stealthily.

2.5 Cyber War

Schneier analyses the strategy of cyber war appropriately, '[a]nd for there to be a cyberwar, there first needs to be a war' [Sb09]. Libicki phrased it similarly arguing that, '[o]perational cyberwar consists of wartime cyber attacks against military targets and military-related civilian targets' [Ml09a]. One of the options for cyber operations is to supplement conventional warfare [CK10] the research refers to this strategy as cyber war. The often hyped 'First Cyberwar' against Estonia was merely a precursor to fully-

fledged cyber war; conducted with low technology means and without any formal declaration of war [Cm07]. At the same time, there were no conventional forms of warfare which those attacks supplemented. If a state of war had been acknowledged by either one or both of the participating states, the operations could have been described as being embedded in a cyber war strategy.

The intensity and objectives with which cyber war can supplement conventional warfare varies. Lonsdale mentioned the ability of cyber warfare to substitute tactical bombing [Ld04]. In general, the intensity of cyber operations during a cyber war is not limited. As Libicki puts it: 'once something is called war, a victim's responsibility for the consequences of its acts dissipates' [Lm96]. Compared to the other kinds of cyber operations, escalation plays a minor role, given that war is already underway. The war can still turn from conventional and cyber weapons to using nuclear weapons (an escalation) but the probability that cyber operations contribute to this escalation rather than conventional warfare is comparatively low. A state would probably more be worried and prone to escalate as response to armies invading its territory and killing its citizens and armies than about the loss of electricity in the capital for example.

The difference between shashou jian and cyber war is not only the setting (cyber war can only take place during war). In addition, cyber war does not necessarily strike stealthily or at decisive points. A cyber war operation could, for example, aim to use distributed denial-of-service attacks to deny the whole country Internet access. It could also utilize destructive viruses to destroy as much data and information within the adversary's state (including private computers, companies etc.) as possible. These broad, destructive and overt operations could be part of a coercing cyber war strategy. They would not fall within a shashou jian framework.

3 Conclusion

The analysis of possible cyber strategies shows that there is a certain cyber strategy for every occasion. In times like this, when the number of stakeholders participating in international cyber conflicts is constantly increasing and no end of hostilities seems likely, it is vital to step up the corresponding security measures, in this case: cyber security.

In terms of strategies, deterrence by denial strategy would be well-chosen to focus on securing the state's 'cyber borders'. This would mean a strong focus, policy- and resource-wise, on enhancing information security – hardening systems, monitoring networks, creating public-private cooperation, research and development e. g. of advanced persistent and volatile threats, sharing information about attacks and raising public awareness. If states wish to engage in offensive cyber strategies, it is even more important to secure the own (critical) national information infrastructure in order to deal with possible retaliation. Thus, improving cyber defenses should always come first.

Cyber security which aims at securing the nation's (critical) information infrastructure

has to take a proactive approach. One way of doing so is to focus research on traps, the so-called honeypots and honeynets – either in a virtual / sandbox environment or as raw steel version. Their research, development and deployment allow the analysis of attack vectors and behaviours, therefore allowing an adaptation of the defensive measures in order to counter future attacks following those patterns. Finding and sharing information about zero-day exploits before their use allows its correction before harm is done. In order to implement a holistic and sustainable cyber security paradigm, knowledge about offensive capabilities is crucial. In cyber security, being seconds too late can already make the difference between having an effective security in place and having none at all.

4 Bibliography

[Aj01] Adams, J.: Virtual Defense. Foreign Affairs 80/3, S. 98-112, 2001.

[BPBF12] Boldizsár, B.; Pék, G.; Buttyán, L.; Félegyházi, M.: The Cousins of Stuxnet: Duqu, Flame, and Gauss, Future Internet 4/6, S. 971-1003, 2012.

[BS11] Betz, D.J.; Stevens, T.: Cyberspace and the State: Toward a Strategy for Cyber-power, The International Institute for Strategic Studies, Routledge, New York, 2011.

[CK10] Clarke, R. A.; Knake, R. K.: Cyber War. The Next Threat To National Security And What To Do About It, Harper-Collings Publisher,^New York, 2010.

[Cm07] Cavelty, M. D.: Critical information infrastructure: vulnerabilities, threats and responses, disarmament forum, ICTs and International Security 3, S. 15-22, 2007.

[Fj08] Fritz, J.: How China will use Cyber Warfare To Leapfrog In Military Competitiveness, Culture Mandala 8/1, S. 28-80, 2008.

[FMC10] Falliere, N.; Murchu, L.O.; Chien, E.: W32.Stuxnet Dossier, Symantec Security Response, Version 1.3, 2010.

[Gb12] Graumann, B.: Cyber-security: The vexed question of global rules, Secure & Defence Agenda, Brussels, 2012.

[Gc99] Gray, C. S.: Modern Strategy, Oxford University Press, Oxford, 1999.

[Gk11] Geers, K.: Strategic Cyber Security, NATO Cooperative Cyber Defence Centre of Excellence, Talinn, 2011.

[Gm11] Gervais, M.: Cyber Attacks and the Laws of War, http://ssrn.com/abstract=1939615, Stand: 15.06.2015.

[Gs61] Snyder, G.: Deterrence and Defense: Toward a Theory of National Security, Princeton University Press, Princeton, 1961.

[He10] Habiger, E. E.: Cyberwarfare and Cyberterrorism: The Need For A New U.S. Strategic Approach, The Cyber Secure Institute, Provoking Cybersecurity Change White Paper Series, White Paper, 2010.

[Hm11] Hjortdal, M.: China's Use of Cyber Warfare: Espionage Meets Strategic Deterrence, Journal of Strategic Security 4/2, S. 1-24, 2011.

[Hs15] Herpig, S.: Anti-War and the Cyber Triangle: Strategic Implications of Cyber Operations and Security for the State, Thesis submitted to the University of Hull for the Degree of Philosophy Doctor (PhD), 2015.

[Ja68] Jomini, A.-H.: The Art of War, J. B. Lippincott & Co., Philadelphia, 1868.

[Ka94] Krepinevich, A .F.: Cavalry to Computer: the Pattern of Military Revolutions, Foreign Affairs 30/13, 1994.

[Kd09] Kuehl, D.T.: From Cyberspace to Cyberpower: Defining the Problem. In (Kramer, F.D.; Starr, S.H.; Wentz, L.K. Hrsg.): Cyberpower and National Security, National Defense University, Washington D.C., S. 24-42, 2009.

[Kf83] Kaplan, F. M.: The Wizards of Armageddon, Simon and Schuster, New York, 1983.

[Kr09] Kugler, R. L.: Deterrence of Cyber Attacks. In (Kramer, F. D.; Starr, S. H.; Wentz, L. K. Hrsg.): Cyberpower and National Security, National Defense University, Washington D. C., S. 309-342, 2009.

[Ld04] Lonsdale, D.: The Nature of War in the Information Age. A Clausewitzian Future, Frank Cass, London, 2004.

[Lj10] Lewis, J. A.: Cross-Domain Deterrence and Credible Threats, Center for Strategic & International Studies, 2010.

[Lm96] Libicki, M. C.: Protecting the United States in Cyberspace. In (Campen, A. D.; Dearth, D. H.; Goodden, T. R. Hrsg.): Cyberwar: Security, Strategy, And Conflict In The Information Age, AFCEA International Press Fairfax, Virginia, S. 91-105, 1996.

[Lm09a] Libicki, M. C.: Cyberdeterrence and Cyberwar, RAND Project AIRFORCE, RAND Corporation, Santa Monica, 2009.

[Lm09b] Libicki, M. C.: Sub Rosa Cyber War. In (Czosseck, C.; Geers, K. Hrsg.): The Virtual Battlefield. Perspectives on Cyber-Warfare 34, IOS Press, Amsterdam, 2009.

[LX10] Lan, T.; Xin, Z., Can Cyber Deterrence Work? In (Nagorski, A. Hrsg.): Global Cyber Deterrence. Views from China, the U.S., Russia, India, and Norway, EastWest Institute, New York, S. 1-3, 2010.

[Nl05] Navrozov, L.: Chinese Geostrategy: 'Assassin's Mace', http://archive.newsmax.com/archives/articles/2005/10/20/172811.shtml, Stand: 21.05.2015.

[PW02] Payne, K. B.; Walton, D. C.: Deterrence in Post Cold-War World. In (Baylis, J.; Wirtz, J.; Cohen, E.; Gray, C. S. Hrsg.): Strategy in the Contemporary World, Oxford University Press, New York, S. 161-182, 2002.

[Rg01] Rattray, G.: Strategic Warfare in Cyberspace, The MIT Press, Cambridge, 2001.

[Rm15] Robinson, M.; Jones, K.; Janicke, H.: Cyber warfare: Issues and challenges, Computers & Security 49, Elsevier, S. 70-94, 2015.

[Sb09] Schneier, B.: So-called Cyberattack Was Overblown, http://www.schneier.com/essay-280.html, Stand 20.04.2011.

[Sa09] Sharma, A.: Cyber Wars: A Paradigm Shift from Means to Ends. In (Czosseck, C.; Geers, K. Hrsg.): The Virtual Battlefield: Perspectives on Cyber-Warfare 34, IOS

Press, Amsterdam, S. 3-17, 2009.

[SG63] Sunzi; Griffith, S. B.: The Art of War, Oxford University Press, Oxford, 1963.

[Ss06] Starr, S. H.: Towards an Evolving Theory of Cyberpower. In (Czosseck, C.; Geers, K. Hrsg.): The Virtual Battlefield: Perspectives on Cyber-Warfare 34, IOS Press, Amsterdam, S. 18-52, 2009.

[Tm67] Tse-Tung, M.: Selected Military Writings of Mao Tse-Tung, Foreign Language Press, Peking, 1967.

[Wi07] Winkler, I.: Zen and the Art of Information Security, Syngress, Rockland, 2007.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

SAG WAS – Studentische AusbildunG und berufliche Weiterbildung in AnforderungsSpezifikationen

SAG WAS – Studentische AusbildunG und berufliche Weiterbildung in AnfordungsSpezifikationen

Karin Vosseberg (Chair)[1]

Die Erhebung, Dokumentation, Abstimmung und Verwaltung von Anforderungen an ein zu entwickelndes System gehört zu den entscheidenden Aufgaben für den gesamten Projekterfolg. Hier werden wesentliche Grundsteine für die Qualität der Software gelegt. Durch Diskussionen über Praktiken wie Behaviour Driven Development (BDD) ist das Thema eng verbunden mit der Softwareentwicklung. Doch wie werden Personen darauf vorbereitet, gute Praktiken des Requirements Engineerings in der Praxis gewinnbringend für die Qualität der Software einzusetzen? Mit Büchern? In Kursen? Learning by doing? Wie wird in der Hochschullehre darauf reagiert? Wie werden dort die Konzepte zur Anforderungsspezifikation vermittelt? In Vorlesungen? Vertieft durch studentische Ausbildungsprojekte? Auch in dem diesjährigen SAG WAS Workshop möchten wir wieder die beiden Welten – studentische Ausbildung und berufliche Weiterbildung zusammenbringen, um einen Austausch zu ermöglichen und voneinander zu lernen.

Für den Workshop haben wir Experten aus dem Bereich Requirements Engineering eingeladen, die mit ihren Vorträgen erste Impulse für die Diskussion setzen. Zusätzlich werden wir im Rahmen einer Postersession unsere Erfahrungen mit unterschiedlichen Veranstaltungskonzepten austauschen.

Programm- und Organisationskomitee:

Prof. Dr. Axel Böttcher, Hochschule München
Dr. Frank Houdek, Daimler AG Ulm
Prof. Dr. Barbara Paech, Universität Heidelberg
Prof. Dr. Ralf Reißing, Hochschule Coburg
Prof. Dr. Andreas Spillner, Hochschule Bremen
Prof. Dr. Veronika Thurner, Hochschule München
Prof. Dr. Karin Vosseberg, Hochschule Bremerhaven (Chair)

URL: www.vosseberg.net/sagwas2015/

[1] Hochschule Bremerhaven, FB 2 – SG Informatik und Wirtschaftsinformatik, An der Karlstadt 8, 27568 Bremerhaven, karin.vosseberg@hs-bremerhaven.de

Workshop-Programm

Wir starten den Workshop in der ersten Sitzung mit *SAG WAS dazu, welche Maßnahmen zur Erstellung von Anforderungsspezifikationen in der Firma eingesetzt und wie die Praktiken an die Mitarbeiter vermittelt werden*. Thomas Garus von der Bredex GmbH wird uns mitnehmen in seinen Projektalltag und die Strategien des Requirements Engineering in den unterschiedlichen Projektkonstellationen diskutieren. Anschließend wird Ulrike Friedrich von Sophist GmbH ihr Konzept von *Just in Time Teaching* als eine Variante von *Blended Learning* in einem betrieblichen Umfeld vorstellen.

In der zweiten Sitzung legen wir den Schwerpunkt auf *SAG WAS dazu, welche Maßnahmen zur Erstellung von Anforderungsspezifikationen in der studentischen Ausbildung wie vermittelt werden*. Barbara Paech von der Universität Heidelberg stellt ihre *Anforderungen an und Erfahrungen zu Requirements Engineering in der universitären Ausbildung* zur Diskussion. Gerd Beneken von der Hochschule Rosenheim nimmt mit *Fail Better: Erfahrungen aus 12 Jahren Requirements Engineering in studentischen Projekten* die Perspektive der Fachhochschulen ein.

In der darauf folgenden Sitzung werden weitere Ausbildungskonzepte in Form von ausgewählten Postern von Veronika Thurner, Axel Böttcher, Kathrin Schlierkamp und Daniela Zehetmeier von der Hochschule München, Yvonne Sedelmaier und Dieter Landes sowie Ralf Reißing, von der Hochschule Coburg und Gero Wedemann von der Fachhochschule Strahlsund vorgestellt und diskutiert. Das Spektrum reicht von Einstiegsprojekten über ein mehrstufiges Lehrkonzept und Blockseminaren bis hin zu weiteren Blended Learning Konzepten.

In der letzten Sitzung kehren wir zurück zur Praxisperspektive mit *SAG WAS dazu, welche Maßnahmen zur Erstellung von Anforderungsspezifikationen in der beruflichen Weiterbildung wie vermittelt werden*. Nicolai Stein von REQUISIS GmbH wird mit seinem Vortrag „RE/RM-Ausbildung im betrieblichen Alltag“ noch mal den Blick auf die berufliche Weiterbildung lenken.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Just-in-Time Teaching - Interne Ausbildung 2.0

Ulrike Friedrich, Karin Schlör[1]

Abstract: Als Wissensvermittler aus dem Bereich Requirements-Engineering arbeiten wir SOPHISTen stetig daran, die Vermittlung von Fachwissen zu verbessern. Dabei entstanden Trainings nach dem Blended Learning Ansatz, bei denen in einer Onlinephase deklaratives Wissen selbst erarbeitet, und in einem abschließenden Workshop auf die Praxis übertragen wird – so wird das sogenannte prozedurale Wissen geschult. Um die interne Ausbildung bei SOPHIST noch effizienter zu gestalten und eine best-mögliche Vorbereitung auf die Praxis sicherzustellen haben wir dieses Konzept weiter verfeinert – das Ergebnis ist Just-in-Time Teaching.

Keywords: Blended Learning, Neue Lernformen, Personalentwicklung, Requirements-Engineering, Anforderungsanalyse, Mitarbeiterbindung, Just-in-Time-Teaching.

1 Die SOPHIST GmbH

Die SOPHIST GmbH ist Anbieter von Beratung und Trainings aus dem Bereich Requirements-Engineering (kurz: RE) unter der fachlichen Leitung von Chris Rupp. Derzeit bieten wir 10 verschiedene Trainings zu 82 festen Terminen 2015 mit unterschiedlichen Schwerpunkten aus dem RE an. Neben der Zertifizierung nach IREB e.V. bieten wir Trainings zu Requirements-Engineering an. Darüber hinaus unterstützen wir mittelständische bis große Unternehmen aus verschiedensten Branchen beim Einsatz von Requirements-Engineering projektbegleitend und durch Coachings.

Um diese Aufgabenbereiche effektiv und für unsere Kunden zufriedenstellend gestalten zu können, benötigen wir fachlich und methodisch bestmöglich geschulte Berater und Trainer. Dies erfordert eine gute interne Aus- bzw. Weiterbildung unserer Mitarbeiter, und als optimales Vorgehen hat sich dabei für uns unter anderem Just-in-Time Teaching (kurz: JiTT) herauskristallisiert.

2 Der Weg zu Just-in-Time Teaching

In den letzten Jahren hat sich die interne Aus- und Weiterbildung in vielen Unternehmen stark verändert. So sind sinnvolle Weiterbildungsangebote nicht mehr nur ein „notwendiges Übel“ das Ressourcen eines Unternehmens verschlingt, sondern ein gewichtiger Faktor einerseits beim Akquirieren neuer Mitarbeiter, andererseits aber auch zur Steigerung der Motivation und Zufriedenheit unter den Kollegen. Wenn es dabei

[1] SOPHIST GmbH, Human Resources, Vordere Cramergasse 13, 90478 Nürnberg, humanresources@sophist.de

auch noch gelingt, fachlich hervorragend ausgebildete Mitarbeiter zu generieren, entsteht eine Win-Win-Situation – denn fachlich und persönlich gestärkte Mitarbeiter verstärken die gesamte Organisation. Wir haben uns deshalb für JiTT, eine Sonderform des Blended Learning, entschieden.

2.1 Blended Learning

Blended Learning ist ein Lernkonzept, das die Vorteile klassischer Lernmethoden und –medien mit den heute verfügbaren Möglichkeiten der Vernetzung über das Internet kombiniert um ein effektives, erfolgreiches und für die Wissensempfänger motivierendes Lernen zu ermöglichen. [Sa04] Die Umsetzung der Wissensvermittlung zeichnet sich durch zwei Phasen aus. So werden in einer Online-Phase Inhalte durch Selbststudium erarbeitet. In einer anschließenden Präsenz-Phase werden diese Inhalte angewendet und vertieft.

2.2 Das didaktische Konzept

Effektive Wissensvermittlung setzt ein sinnvolles didaktisches Konzept voraus. Ein Grundpfeiler unseres Trainings bildet die konstruktivistische Didaktik, die einen „Wechsel vom Lehren zum Lernen“ an-strebt. Um die Selbststeuerung des Lernprozesses zu gewährleisten, ist es die Aufgabe des Lehrenden, optimale Lernbedingungen zu schaffen statt Wissen zu übertragen. [Mo09]

Der zweite wesentliche Aspekt, den es bei der Vermittlung von Wissen zu beachten gilt, ist die Unterscheidung zwischen deklarativem und prozeduralem Wissen. Deklaratives Wissen bezeichnet das theoretische Wissen, also Faktenwissen. Diese Art des Wissens ist verhältnismäßig leicht zu vermitteln. Im Gegensatz dazu gibt es das prozedurale Wissen, das beschreibt wie man Wissen in der Praxis anwendet. Dieses prozedurale Wissen zu vermitteln ist die größte Schwierigkeit, denn zwischen Verstehen und Anwenden können liegt viel Arbeit.

Gerade im Requirements-Engineering ist es essentiell, zunächst einmal den theoretischen Hintergrund, beispielsweise aus den Bereichen der Ermittlung und Dokumentation von Anforderung zu verstehen, bevor das Hauptziel, Techniken aus dem RE in der Praxis einzusetzen, angegangen werden kann.

Blended Learning stellt sich dabei als der optimale Ansatz heraus, da die grundlegenden Theorien in der Online-Phase erarbeitet, und im Anschluss daran in einer Präsenzveranstaltung, unter Anleitung erfahrener Trainer, angewendet werden können. Durch den Einsatz einer Lernplattform, auch Learn Management System, und dem damit verbundenen Aufbereiten und Bereitstellen von Lerninhalten sowie Selbsttests zur Überprüfung des Lernfortschritts wird darüber hinaus eine Lehre nach der konstruktivistischen Didaktik während der Online-Phase angewendet.

Blended Learning erlaubt das Vermitteln von prozeduralem Wissen indem der Erwerb von deklarativem Wissen in eine Online-Phase ausgelagert wird. Dadurch kann die benötigte Zeit zum Anwenden von Wissen geschaffen werden – Zeit, die bei konventionellem Vorgehen sonst leider oftmals auf der Strecke bleibt.

Der Blended Learning Ansatz lässt viele Spielräume, wie Online- und Präsenz-Phasen angeordnet werden. Im Rahmen unserer Blended Learning Trainings haben wir uns dazu entschieden, zunächst mit einer Online-Phase zu starten. Jeder Teilnehmer muss zu jedem relevanten Themenbereich einen Test bearbeiten und diesen auch, mit mindestens 60% aller erreichbaren Punkte, bestehen. Nur so kann ein Grundniveau an deklarativem Wissen bei allen Teilnehmern zu Beginn der Präsenzveranstaltung gesichert werden – denn dieses Vorwissen ist zwingend notwendig um bei allen Übungen erfolgreich teilnehmen, und so prozedurales Wissen erlernen zu können.

2.3 Just-in-Time Teaching

Just-in-Time Teaching als Sonderform des Blended Learning zeichnet sich durch ein iteratives Vorgehen aus, das auf einer kurzfristigen Vorbereitung der Präsenzphase anhand der Ergebnisse aus der Online-Phase beruht. Die Präsenzphase wird also „just in time“ vorbereitet.

Der Ablauf

Im wöchentlichen Rhythmus bereiten sich die Teilnehmer unseres Just-in-Time Teaching Programms zunächst selbstständig vor. Sie bekommen dazu Lerninhalte zu einem abgegrenzten Themengebiet aufbereitet und auf dem SOPHIST Learn Management System zur Verfügung gestellt. Zu jedem Themengebiet gibt es auch hier einen automatisiert auswertbaren Test.

Die Teilnehmer eignen sich die Lerninhalte an und beantworten im Anschluss den Selbsttest. Durch die automatisiert auswertbaren Antworten erhalten sie ein unmittelbares Feedback über den Lernerfolg und können so bei Bedarf gezielt Teilgebiete wiederholen und vertiefen. So kann deklaratives Wissen effektiv gelernt werden. Darüber hinaus fertigt jeder Teilnehmer jede Woche einen Learning Log an.

Der Learning Log

Ein Learning Log, auch Lerntagebuch, unterstützt den Lernerfolg indem Leitfragen beantwortet werden, die zum Reflektieren der Lerninhalte anregen.

Die Präsenz-Veranstaltung

Für die Vorbereitung der Präsenz-Veranstaltung wertet der zuständige Trainer die Ergebnisse aus den Online-Tests sowie den Learning Log aus. Je nach Testergebnissen und aufgeworfenen Fragen wird das Konzept der Präsenzveranstaltung erstellt. Diese besteht in der Regel aus zwei Teilen. So werden zunächst offene Fragen aus den

Learning Logs diskutiert und beantwortet, sowie schwierige Fragen aus den Selbsttests erneut aufgegriffen und besprochen. Anschließend wird das prozedurale Wissen geschult indem speziell auf die Themenbereiche zugeschnittene Übungen durchgeführt werden. In diesen Übungen ist eine hohe Aktivität der Beteiligten notwendig, denn nur durch das eigenständige, wenn auch angeleitete, Anwenden von Wissen kann prozedurales Wissen erlangt werden.

3 Fazit

Die Vermittlung von Wissen ist ein breites Feld, auf dem noch großer Forschungsbedarf besteht. Es gibt kein Allheilmittel für die bestmögliche Wissensvermittlung, doch es ist immer sinnvoll, ein didaktisches Konzept als Grundlage zu verwenden. Durch das Verschmelzen unserer langjährigen Praxiserfahrung mit neuesten wissenschaftlichen Erkenntnissen aus der Kooperation mit Hochschulen schaffen wir stetig neue Ansätze zur Vermittlung von Wissen und beschäftigen uns intensiv mit neuen Vermittlungsmethoden.

Unsere Erfahrungen haben gezeigt, dass Just-in-Time Teaching als Sonderform des Blended Learning ein hervorragendes Vorgehen für die interne Weiterbildung darstellt, da so genügend Zeit für das An-wenden von Wissen in einer Präsenz-Veranstaltung geschaffen werden kann. Darüber hinaus bietet JiTT die Möglichkeit, den Lernprozess individuell auf die Bedürfnisse der Lernenden zuzuschneiden und dadurch ein bestmögliches Ergebnis bei der Wissensvermittlung zu erzielen.

Literaturverzeichnis

[Fi13] Figas P., Hagel G., Bartel A., The Furtherance of Motivation in the Context of Teaching Software Engineering. In: Institute of Electrical and Electronics Engineers (Hg.) Global Engineering Education Conference EDUCON, Berlin, S.1299–1304, 2013.

[Ha13] Hagel G., Mottok J., Müller-Amthor M., Drei Feedback-Zyklen in der Software Engineering-Ausbildung durch erweitertes Just-in-Time-Teaching, In: Software Engineering im Unterricht der Hochschulen, SEUH 2013.

[Mo09] Mottok J., Hagel G., Utesch M., Waldherr F-. Konstruktivistische Didaktik -ein Rezept für eine bessere Software Engineering Ausbildung? In: Proceedings of the 2nd Embedded Software Engineering Conference, ISBN 978-3-8343-2402-3, S. 601-610, December 2009.

[Sa04] Sauter A., Sauter M., Werner; Bender H.: Blended Learning: Effiziente Integration von E-Learning und Präsenztraining, 2. Auflage, 2004, Neuwied: Hermann Luchterhand.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Anforderungen an Anforderungen — Vermitteln von Requirements Engineering

Veronika Thurner[1] Axel Böttcher[1] Kathrin Schlierkamp[1] Daniela Zehetmeier[1]

Abstract: Ungeklärte oder nur scheinbar geklärte Anforderungen sind ein erhebliches Risiko für den Erfolg von Softwareprojekten. Entsprechend müssen angehende Softwerker/innen im Studium gute Praktiken des Requirements Engineerings erlernen. Dazu ist es notwendig, frühzeitig das Bewusstsein für die Bedeutung von Anforderungen und für Qualitätsmerkmale guter Anforderungsdefinitionen zu entwickeln. Ein haptisch angelegtes Einstiegsprojekt leistet dazu einen ersten Beitrag.

Keywords: Requirements Engineering, Einstiegsprojekt, Didaktik

1 Motivation und Grundproblem

Viele Informatik-nahe Studiengänge fokussieren in der Grundausbildung zunächst die Softwareentwicklung. Nachdem die implementierungstechnischen Grundlagen geschaffen wurden setzen sich die bereits etwas fortgeschritteneren Studierenden dann in den mittleren Semestern ihres Studiums intensiv mit Software und Requirements Engineering auseinander. In der studentischen Wahrnehmung ist damit zunächst das Implementieren als zentraler Aufgabenbereich verankert, in der Regel assoziiert mit einer gewissen Technik-Orientierung, klaren Vorgaben, einem hohen Genauigkeitsanspruch und der eindeutigen Entscheidbarkeit, ob eine Lösung richtig ist oder nicht.

Das Themenfeld der Anforderungen erscheint vielen Studierenden daraufhin zunächst erst einmal schwammig und vergleichsweise ungenau zu sein. Gleichzeitig sind die Bedeutung und der Nutzen der Anforderungsdefinition zunächst nur bedingt ersichtlich, da die erstellte Dokumentation (anders als der in den früheren Semestern erstellte Quelltext) auf den ersten Blick keinen unmittelbar erkennbaren Beitrag zum lauffähigen Produkt leistet.

2 Zielsetzung

Um bereits sehr frühzeitig im Studium ein Bewusstsein für die zentrale Bedeutung von Anforderungen und systematischem Requirements Engineering zu entwickeln haben wir ein Lehrkonzept für ein Einstiegsprojekt entwickelt, das die Studierenden bereits unmittelbar zu Beginn ihres Studiums an die Kerntätigkeiten des Software Life Cycles und damit auch an das Thema *Anforderungen* hinführt. Unser Ziel ist es, die Studierenden aktiv erfahren zu lassen, dass auch auf den ersten Blick scheinbar einfache Aufgabenstellungen der präzisen Klärung bedürfen. Des Weiteren werden erste Schlüsselfragen und Vorgehensschritte eingeführt, die zu einer systematischen Klärung der Anforderungen beitragen.

3 Vorgehensweise

Dieses Einstiegsprojekt wird an der Fakultät für Informatik und Mathematik der Hochschule München im Rahmen der Studieneinführungstage durchgeführt. Diese dienen einerseits der ersten Orientierung und sozialen Vernetzung der Studierenden untereinander.

[1] Hochschule München, Fakultät für Informatik und Mathematik, Lothstraße 64, 80335 München, <vorname>.<nachname>@hm.edu; Gefördert durch das BMBF Förderkennzeichen 01PL11025 (Projekt „Für die Zukunft gerüstet“), im Programm „Qualitätspakt Lehre“

Zum anderen schaffen sie einen schnellen Einstieg in den Prozess des Studierens an sich, sowie in das gewählte Studienfach, in unserem Falle Informatik, Wirtschaftsinformatik, Scientific Computing oder Geotelematik.

Als Aufgabe wählten wir das Sortieren eines Kartenstapels, also eine Aufgabenstellung, die den Teilnehmenden zunächst intuitiv verständlich erschien. Die Studierenden arbeiteten dabei in kleinen Gruppen von bis zu 10 Personen. Jeder Gruppe war eine Betreuungsperson zur Seite gestellt, die unter anderem die Rollen *Kunde* und *Tester/in* übernahm.

Als haptisches Arbeitsmaterial stand jeder Gruppe zunächst ein Kartenstapel zur Verfügung, dessen Karten duplikatfrei mit natürlichen Zahlen nummeriert waren. Auf dieser Grundlage beschrieben die Studierenden in natürlicher Sprache ein Sortierverfahren, das sie zuvor anhand des initialen Materials erarbeitet hatten. In dieser Phase erschienen die Anforderungen aus der Aufgabenstellung noch sonnenklar und eher einfach zu sein.

Im nächsten Schritt wurden Testfälle ausgegeben, anhand derer die zuvor erarbeitete Lösung jeweils zu erproben war. Ein Testfall bestand aus Karten mit Buchstaben, ein weiterer aus nicht fortlaufend nummerierten Karten. Schafkopfkarten waren ebenso vertreten wie der leere Kartenstapel und ein Stapel mit Bilderkarten. Schnell wurde deutlich, dass die erstellten Lösungen sehr stark auf das initiale Beispiel ausgerichtet waren, einer allgemeineren Sortieraufgabe jedoch nicht mehr gerecht wurden.

Darauf aufbauend wurden Schlüsselfragen erarbeitet, die das methodische Klären der Anforderungen erleichtern. Unter anderem wurde festgelegt, was die Eigenschaft des Sortiertseins eigentlich ausmacht und dokumentiert, welche Voraussetzungen erfüllt sein müssen, damit ein Kartenstapel überhaupt sortierbar ist. Des Weiteren wurden Randfälle identifiziert sowie geeignete Testfälle ermittelt, anhand derer die Lösung gegen die definierten Anforderungen gestestet werden kann. Ebenfalls diskutiert wurden Qualitätskriterien für Anforderungen, wie beispielsweise Eindeutigkeit, Verständlichkeit und Überprüfbarkeit.

4 Erfahrungsbericht und Lessons Learned

Die Aufgabe des Sortierens erschien vielen Studierenden zunächst eher trivial zu sein, waren sie doch alle aus ihrem Alltagsleben mit einer derartigen Aufgabenstellung vertraut. Schnell wurde jedoch deutlich, dass es innerhalb der Gruppe durchaus unterschiedliche Sichtweisen auf die Anforderungen gab (beispielsweise hinsichtlich der Sortierrichtung oder der überhaupt anzuwendenden Ordnung als Sortierkriterium), obwohl zunächst alle Gruppenmitglieder bestätigt hatten, dass ihnen die Aufgabe klar ist. Auch die Möglichkeit eines leeren oder einelementigen Kartenstapels wurde von den Gruppen im Vorfeld nicht bedacht, sodass die initial gefundenen Lösungen schnell an ihre Grenzen stießen.

So wurde anschaulich erfahrbar, dass selbst bei dieser einfachen, verstanden geglaubten Aufgabenstellung das genaue Klären der Anforderungen erstens überhaupt notwendig ist; zweitens auch einen nicht-trivialen Prozess erfordert, um zu einem zufrieden stellenden Ergebnis zu kommen; und dass drittens für die präzise, für andere nachvollziehbare Dokumentation der erkannten Anforderungen Techniken benötigt werden, die weit über die Verwendung natürlicher Sprache hinaus gehen. All dies ist notwendig, um die Vorasussetzungen zu schaffen für eine spätere Lösung und diese so überhaupt erst zu ermöglichen.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Ein mehrstufiges Lehrkonzept für RE

Yvonne Sedelmaier[1], Dieter Landes[1]

Abstract: Der Beitrag stellt ein didaktisches Konzept vor, das darauf abzielt, Studierenden die Bedeutung von Anforderungen und die Herausforderungen bei deren Erhebung und Dokumentation aufzuzeigen, bevor es mittels aktivierender Lehrmethoden Lösungen dafür anbietet. Dazu versetzen sich Studierende u.a. in mögliche Kunden hinein, bevor sie ein Kundengespräch mit einem echten Stakeholder vorbereiten, durchführen und auf dieser Grundlage ein Anforderungsdokument verfassen.

Keywords: aktive Lehrformen; induktive Lehrformen; Didaktik; Anforderungserhebung; reale Stakeholder

1 Einleitung

Anforderungen entscheiden über den Projekterfolg. Dennoch ist es eine Herausforderung, RE zu lehren: Da theoretisches Vorwissen erforderlich ist, können praktische RE-Projekte erst gegen Ende des Studiums stattfinden. Studierenden fehlt häufig das Problembewusstsein. Traditionelle Hochschulausbildung im RE bietet Studierenden oft Lösungen für Probleme, die sie noch nicht einmal erahnen können. Denn bisher erhielten sie meist kleine Programmieraufgaben, die auf eine spezifische, isolierte, präzise und für Informatikstudierende verständlich formulierte Problemstellung fokussieren. Im Berufsalltag kommen Anforderungen aber meist von Stakeholdern, die sie nicht auf einem Silbertablett servieren und meist mit Fachbegriffen der Informatik wenig verbinden. Studierende glauben häufig, Anforderungen wären einfach zu erheben und zu dokumentieren und unterschätzen so die Bedeutung überfachlicher Kompetenzen im RE.

2 Mehrstufiges didaktisches Konzept

Das beschriebene didaktische Konzept zielt auf ein Wahlpflichtfach im 4. Semester des Bachelorstudiengangs Informatik an der HAW Coburg mit ca. 20 Studierenden. Als Lehrziele sollen Studierende neben Fachwissen wie etwa Erhebungstechniken vor allem praktisches Handlungswissen und Erfahrungen zum Thema RE sammeln, die Bedeutung von Kommunikation und weiterer kontextsensitiver überfachlicher Kompetenzen [SL15] erkennen sowie ihre Selbstreflexion und Eigenverantwortung verbessern.

Die Studierenden bringen bereits theoretisches Vorwissen zum Thema Anforderungs-

[1] Hochschule für Angewandte Wissenschaften Coburg, Fakultät Elektrotechnik und Informatik, Friedrich-Streib-Str. 2,96450 Coburg, {yvonne.sedelmaier, dieter.landes}@hs-coburg.de

spezifikation mit Use Cases mit. In einer ersten Aufgabe sollen sie sich ihr Vorgehen überlegen, um ein Angebot für eine Seminarverwaltungssoftware zu erstellen. Dazu sollen sie ihre nächsten Schritte begründen und mögliche Schwierigkeiten bei der Angebotserstellung aufzeigen. Im Dialog wird dann gemeinsam erarbeitet, welche Kriterien ein Kundengespräch professionell wirken lassen. Die Studierenden wechseln die Perspektive und verstehen so, worauf potentielle Kunden achten. Im Anschluss werden zentrale Elemente eines solchen Gespräches zur Anforderungserhebung erarbeitet. So wird bei den Studierenden zunächst ein Bewusstsein für mögliche Problemstellungen geweckt, eine zentrale Voraussetzung für erfolgreiches Lernen [Si99].

Je 4-5 Studierende bereiten dann als Gruppe ein tatsächliches Kundengespräch vor und berücksichtigen dabei die theoretisch bekannten Aspekte. Dann führen sie ein einstündiges Gespräch mit einem echten Stakeholder. In der nächsten Lehrveranstaltung erhalten die Studierenden Videoaufzeichnungen ihres Gesprächs und einige Leitfragen, um das Gespräch in der Gruppe, mit den Lehrenden und der Kundin zu reflektieren. Diese Gespräche bilden die Basis für das Erstellen eines echten Anforderungsdokuments: Die Ergebnisse werden strukturiert, als Prozessmodell dokumentiert, gereviewt, Anforderungen abgeleitet und in einem Anforderungsdokument festgehalten.

3 Evaluation und Ausblick

Aus Sicht der Lehrenden wie auch der Studierenden unterstützen solche induktiven und aktivierenden Lehrmethoden dabei, die Bedeutung von Anforderungen und Herausforderungen bei deren Erhebung besser zu verstehen. So ist etwa den meisten Studierenden klar, was theoretisch zu einer positive Gesprächsatmosphäre im Kundengespräch führt. Trotzdem gelingt es manchmal nicht, eine solche Atmosphäre herzustellen. Die Studierenden sehen nun aber für sich einen deutlichen Kompetenzzuwachs in den adressierten überfachlichen Kompetenzen [SL14]. Daher soll das didaktische Konzept künftig noch verfeinert werden.

Literaturverzeichnis

[Si99] Siebert, H.: Pädagogischer Konstruktivismus. Eine Bilanz der Konstruktivismusdiskussion für die Bildungspraxis. Luchterhand, Neuwied, Kriftel, 1999.

[SL14] Sedelmaier, Y.; Landes, D.: A Multi-Level Didactical Approach to Build up Competencies in Requirements Engineering. In (Penzenstadler, B.; Gregory, S.; Landes, D. Hrsg.): 8th International Workshop on Requirements Engineering Education & Training (REET 2014). CEUR Workshop Proceedings vol. 1217; S. 26–34, 2014.

[SL15] Sedelmaier, Y.; Landes, D.: Überfachliche Kompetenz im Software Engineering - Modellierung, Förderung und Messung in der Hochschulausbildung. In (Riegel, U. et al. Hrsg.): Kompetenzmodellierung und -messung in den Fachdidaktiken. Waxmann, Münster; S. 111–130, 2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Requirements Engineering an der Hochschule in einer Blockwoche – Best Practices und Lessons Learned

Gero Wedemann[1]

Abstract: An der Fachhochschule Stralsund wird Requirements Engineering seit einigen Jahren in einer viertägigen Blockwoche gelehrt. Verschiedene Varianten der Vermittlung wurden dabei erprobt. In diesem Artikel werden die aus unserer Sicht wichtigsten Faktoren für Erfolg und Misserfolg dargestellt.

Keywords: Requirements Engineering, Blockveranstaltung

1 Einleitung

Die Lehrveranstaltung „Requirements Engineering“ ist eine Pflichtveranstaltung in den Bachelor-Studiengängen „Angewandte Informatik“ und „Medizininformatik und Biomedizintechnik“ am Anfang des 4. Semesters sowie im Master-Studiengang Elektrotechnik. Die insgesamt ca. 50 Teilnehmer werden von einem Professor und zwei Mitarbeitern betreut. Bei Planung und Durchführung dieser Lehrveranstaltung haben wir in den vergangenen Jahren viel experimentiert. Seit dem letzten Durchgang sind wir zum ersten Mal mit dem Konzept hinreichend zufrieden und haben auch positives Feedback der Studierenden erhalten. In diesem Artikel wird dargestellt, was sich im Verlauf der letzten Jahre als besonders nützlich erwiesen hat („Best Practice“) und was als hinderlich („Lessons Learned“). Wir sind sicher, dass sich diese Erfahrungen auch auf andere Kurse zu diesem Thema übertragen lassen.

2 Best Practices

Der Kurs beginnt bei der Identifizierung der Stakeholder und der Opportunity, der Summe der Bedürfnisse der Stakeholder [Es14], und führt bis zur Entwicklung eines Papier-Prototyps [Kl14]. Die Inhalte werden nicht nur theoretisch erklärt und anhand kleiner Übungsaufgaben demonstriert, sondern die Studierenden entwickeln in Übungen auch ein überschaubares aber hinreichend komplexes Beispiel. Die Entwicklung bis hin zum Papierprototyp hat sich als hilfreich erwiesen, da die Studierenden nicht nur die Spezifikation sondern auch am Ende eine (fast) vollständige Applikation erarbeitet haben und so den Gesamtkontext der Aktivitäten besser verstehen. Das durchlaufende Beispiel sollte es aus einer Domäne stammen, die die Studierenden verstehen und zu der

[1] Fachhochschule Stralsund, Fachbereich Elektrotechnik und Informatik, Zur Schwedenschanze 15, 18435 Stralsund, gero.wedemann@fh-stralsund.de

die Betreuer qualifiziert Auskunft geben können. In unserem Kurs entwickeln wir eine Anwendung für den Antrag auf Urlaub für Mitarbeiter der FH. An diesem Beispiel lassen sich auch dynamische und statische Modelle gut erarbeiten. Der aktuelle Kurs findet an vier vollen Tagen statt, an denen etwa 4 Einheiten von 90 Minuten als in der Summe 16 Einheiten gearbeitet wird. Kürzere Kurse haben sich nicht bewährt, da dann nicht genügend Zeit ist, den gesamten Zyklus praktisch und in der notwendigen Tiefe zu behandeln. Die Vorlesungen zur Vermittlung der Inhalte wurden stark auf das Notwendigste eingekürzt, so dass pro Tag nur zwei bis drei interaktive Vorlesungen von 45 Minuten stattfinden. Die restliche Zeit wird in kleineren Gruppen gearbeitet und beinhaltet zu einem kleineren Teil Übungen und einem größeren Teil die Bearbeitung des durchlaufenden Beispiels. Bei der Bearbeitung der Aufgaben eignen sich die Studierenden notwendige Details der Methoden selbständig anhand von gestellten Unterlagen an. Essentiell ist eine gute Abstimmung der beteiligten Dozenten über Inhalt aber auch den Gruppenprozess in täglichen Abschlussbesprechungen.

Im Rahmen der Veranstaltung werden die folgenden Methoden eingeführt, die sich besonders für diesen Rahmen eignen: Stakeholder-Matrix, Vision, Storyboard, User Storys, Qualitätsmerkmale nach ISO 25010, UML: Zustands-, Aktivitäts-, Klassendiagramm, ausgewählte Analysemuster, Dialogablaufplan, Papierprototyp. Weitere Methoden werden in den vorlesungsartigen Teilen kurz vorgestellt und diskutiert, nachdem die Studierenden praktische Erfahrungen sammeln konnten.

3 Lessons Learned

Umfangreicher Input im Sinne von Vorlesungen hat sich nicht bewährt, da daraus die Studierenden in der Regel wenig mitnehmen und die Übungen frustrierend für alle Seiten sind. Für das durchlaufende Beispiel ist es wichtig, kein zu komplexes Beispiel zu wählen, das in der vorhandenen Zeit zu machen ist, da dies die Studierenden unbefriedigend empfinden. Zu einfache Beispiele sind aber auch nicht zielführend, da sich der Sinn der Methoden daran nicht erschließt. Wir verzichten inzwischen auf das Durcharbeiten größerer Vorlagen wie z.B. für Vision und Scope, da dies für die Studierenden langweilig ist, wenig Neues bringt und so eher zu einer Ablehnung dieser Methoden führt. Vorlesungen, die neuen Stoff beinhalten, sollten nicht in der letzten Einheit quasi auf Vorrat gehalten werden, da die Studierenden in der Regel dann dafür zu erschöpft sind. Stattdessen bieten sich hier Überblicke und Diskussionen an.

Literaturverzeichnis

[Es14] Kernel and Language for Software Engineering Methods (Essence). Version 1.0. Object Management Group, 2014.

[Kl14] Klemmer, S.: Human-Computer Interaction: 3.1 Storyboards, Paper Prototypes, and Mockups. Coursera, https://www.coursera.org/course/hciucsd, 2014.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Lehren und Prüfen im Requirements Engineering am Beispiel eines Master-Kurses für Ingenieure

Ralf Reißing[1]

Abstract: Requirements Engineering (RE) ist eine der wichtigsten Aktivitäten in Entwicklungsprojekten, denn hier entscheidet sich der Projekterfolg in großem Maße. Die RE-Ausbildung an Hochschulen ist daher notwendig, darf allerdings nicht nur aus der Vermittlung von Faktenwissen bestehen. Entscheidend für den Lernerfolg sind vielmehr das Reflektieren und das praktische Tun. In diesem Beitrag wird ein Ansatz vorgestellt, mit Hilfe von Just-in-Time-Teaching und Projektarbeit in realen Projekten RE-Kompetenz bei Studierenden in einem Ingenieurstudiengang aufzubauen, zu vertiefen und kompetenzorientiert zu prüfen.

Keywords: Requirements Engineering, Just-in-Time-Teaching, kompetenzorientiertes Prüfen

1 Einleitung

Der Master-Studiengang Entwicklung und Management im Maschinen- und Automobilbau der Hochschule Coburg bietet ein großes Wahlpflichtangebot. Das Modul Requirements Engineering und Management (REM; 5 ECTS) vermittelt RE für technische Systeme für (Wirtschafts-)Ingenieure. Die Teilnehmer sollen lernen, im Team projektspezifisch ein geeignetes RE-Vorgehen zu erarbeiten sowie möglichst vollständige, verständliche und abgestimmte Anforderungsspezifikationen zu liefern.

2 Lehren und Prüfen

Grundlage der Lehre ist das Buch Requirements Engineering und -Management [Ru09], das umfangreiches Erfahrungswissen einer großen RE-Beratungsgesellschaft bündelt. Daraus werden die Anteile ausgewählt, die für das RE technischer Systeme besonders relevant sind. Als Einstieg dient ein Artikel von Ebert [Eb12], der das Thema in seiner Gesamtheit an einem technischen Beispiel mit Tipps für die Praxis darstellt.

Die Lehrform im Modul ist Just-in-Time-Teaching (JITT), orientiert an der Umsetzung für RE durch Hagel et al. [HMM13]. JITT verlagert die Erarbeitung des Stoffs in die Selbstlernphase, während die Präsenzveranstaltung für die Diskussion offener Fragen, Beispiele und Übungen genutzt wird. Um das Selbstlernen zu unterstützen, bekommen die Teilnehmer ergänzend zur Angabe des zu lesenden Kapitels Aufgaben zum Stoff, die sie selbstständig bearbeiten, bspw. ein Zustandsdiagramm zu modellieren. Ihre

[1] Hochschule Coburg, Fak. Maschinenbau und Automobiltechnik, 96450 Coburg, ralf.reissing@hs-coburg.de

Ergebnisse geben Sie über das E-Learning-System Moodle ab. Dazu gehören auch Unklarheiten und offenen Fragen sowie Vorschläge für Prüfungsaufgaben. Diese Abgaben dienen dem Dozenten als Basis für die Vorbereitung und Gestaltung der Präsenzveranstaltung, in der auch einzelne Werkzeuge zur UML-Modellierung (z.B. Enterprise Architect) und zum Requirements Management (z.B. DOORS) angewendet werden. Ergänzend dazu führen die Studierenden selbstständig ein semesterbegleitendes RE-Projekt im Umfang von 1-2 SWS mit echten Stakeholdern durch, an dem das erarbeitete Wissen und die Werkzeuge praktisch angewendet und vertieft werden. Der Dozent steht dem Projekt bei Bedarf als Coach zur Verfügung.

Die Bewertung des Lernerfolgs der Studierenden erfolgt nicht durch eine schriftliche Prüfung, sondern anhand der Projektergebnisse. Das ist zum einen ein sog. Wissenschaftlicher Bericht, in dem die Studierenden ihre Vorgehensweise im RE-Projekt (das RE-Konzept) beschreiben, begründen und reflektieren (Erfolge, Misserfolge, Lessons Learned). Hier kommt es auf zielführende, projektangemessene und gut begründete Entscheidungen an. Zum anderen geht die Qualität der im Projekt erstellten Anforderungsspezifikation in die Benotung ein.

3 Erfahrungen

Das Modul REM fand bisher zwei Mal statt. Durch die kleine Teilnehmerzahl von vier bzw. fünf Studierenden gelang in der Präsenzveranstaltung dank der Vorbereitung in der Selbstlernphase eine intensive Diskussion und Übung. Offene Fragen konnten geklärt und Bezüge zur gelebten Praxis des RE für technische Systeme hergestellt werden. In der Projektarbeit[2] konnten typische Erfahrungen wie nicht verfügbare Stakeholder, viele offene Fragen oder widersprüchliche Anforderungen gemacht werden. Insgesamt gab es eine effektive Projektarbeit mit guten Ergebnissen.

Literaturverzeichnis

[Eb12] Ebert, C.: Anforderungen ermitteln: Zehn Schritte für die Praxis. OBJEKTspektrum, 06/2012.

[HMM13] Hagel, G.; Mottok., J.; Müller-Amthor, M.: Drei Feedback-Zyklen in der Software Engineering-Ausbildung durch erweitertes Just-in-Time-Teaching. In (Spillner, A.; Lichter, H. Hrsg.): Software Engineering im Unterricht der Hochschulen (SEUH) 2013. LNI 956, Bonn, S. 17-26, 2013.

[Ru09] Rupp, C..: Requirements Engineering und -Management. Hanser Verlag, 2009.

[2] Projektgegenstand 2014: ein Verstellsystem für die Heckflügel eines Formula-Student-Fahrzeugs in Kooperation mit dem CAT-Racing-Team der Hochschule. 2015: ein Messsystem für den Überholabstand von Fahrradüberholern in Kooperation mit dem ADFC Coburg, der Polizei und der Stadt Coburg.

Hochschule 2025

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Hochschule 2025

Djamshid Tavangarian[1], Ulrike Lucke[2], Markus von der Heyde[3]

Die wachsende Bedeutung der Mikro- und Makromobilität von Studierenden, Mitarbeitern und Geräten, allgegenwärtigem IT-Zugang, dynamischer Allokation von personalisierten Inhalten sowie flexibler Gestaltung und Organisation von Community- und individuellen Lern- und Arbeitsprozessen stellt moderne Bildungseinrichtungen bei ihrer stetigen Entwicklung vor technische, organisatorische und inhaltliche Herausforderungen. Der Workshop thematisiert diese Herausforderungen und stellt sie zur Diskussion. Dabei liegt der Fokus 2015 auf den besonderen Herausforderungen IT-Governance, Campus Management und der Datenanalyse als Basis von Entscheidungsprozessen.

Aus den zwölf eingereichten Beiträgen wurden durch ein Peer-Review neun Beiträge zur Präsentation auf dem Workshop ausgewählt. Abgerundet werden diese Beiträge durch eine Keynote zum Thema Projektportfoliomanagement an Hochschulen sowie einen eingeladenen Vortrag zum Thema Campus Management aus Sicht eines Software-Anbieters.

Wir danken den GI-Fachgruppen Arbeitsplatzrechensysteme und Personalcomputer (APS+PC) sowie E-Learning (ELE) für die Unterstützung bei der Realisierung des Workshops. Für die sorgfältige und engagierte Unterstützung bei der Organisation des Workshops danken wir ganz herzlich Raphael Zender und Francis Zinke. Zudem gebührt den Mitgliedern des Programmkomitees für die Unterstützung bei der Begutachtung der eingereichten Beiträge unser Dank:

Arndt Bode (LRZ München)
Michael Breitner (Uni Hannover)
Christian Erfurth (FH Jena)
Stefan Fischer (Uni Lübeck)
Marc Göcks (Multimedia Kontor Hamburg)
Heinz-Gerd Hegering (LMU München)
Christoph Igel (Uni des Saarlandes)
Odej Kao (TU Berlin)
May-Britt Kallenrode (Uni Osnabrück)
Udo Kebschull (Uni Frankfurt)
Frank Klapper (Uni Bielefeld)
Michael Koch (UniBW München)
Bernd Krämer (FU Hagen)
Gudrun Oevel (Uni Paderborn)
Hans Pongratz (TU München)
Christa Radloff (Uni Rostock)
Peter Schirmbacher (HU Berlin)
Ulrik Schroeder (RWTH Aachen)
Bernd E. Wolfinger (Uni Hamburg)
Markus Zahn (Uni Augsburg)

[1] Universität Rostock, Institut für Informatik, A.-Einstein-Straße 22, 18059 Rostock, djamshid.tavangarian@uni-rostock.de
[2] Universität Potsdam, Institut für Informatik, A.-Bebel-Str. 89, 14482 Potsdam, ulrike.lucke@uni-potsdam.de
[3] vdH-IT, Rosenweg 7, 99425 Weimar, info@vdh-it.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Projektflut - geht der CIO baden? Multiprojektmanagement an Hochschulen

Sören Lorenz [1], Frank Klapper [2]

Abstract: Die kontinuierliche Zunahme von Projekten mit IT-Bezug in Zeiten der digitalen Transformation stellt die Hochschulen vor große organisatorische Herausforderungen. Projektarbeit etabliert sich mehr und mehr als Daueraufgabe in den Servicebereichen einer Hochschule. Sie bindet qualifiziertes Personal und belastet die Linienorganisation. Multiprojektmanagement als Methode kann hier Lösungsperspektiven aufzeigen. Strukturell etabliert in einem Projektmanagementoffice (PMO) können Projekte auf einer strategischen Ebene koordiniert und Projektmanagementwissen effektiv in der Organisation verankert werden. Als Serviceeinheit kann das PMO zudem als Steuerungsmodell für die Stabilisierung von Matrixstrukturen verstanden werden, da stabile Matrixstrukturen die Voraussetzung für erfolgreiche Projektarbeit in *cross-functional* Teams sind.

Keywords: Multiprojektmanagement, Projektportfoliomanagement, Matrixorganisation, cross-functional Teams, Organisationsentwicklung, IT-Projekte, CIO.

1 Ausgangslage

Die Einführung von IT- und Informationssystemen in Hochschulen bestimmen seit mehr als einer Dekade die Diskussion auf zahlreichen Tagungen und in Dachorganisationen (siehe insb. ZKI-Tagungen und Arbeitsgruppen[3] und DINI-Tagungen und Workshops[4]). Die Themen reichen in alle Kernbereiche der Hochschulen und betreffen Lehre, Forschung und insbesondere Service[5]. IT- und Informationssysteme verändern den Arbeitsalltag und entwickeln sich als Treiber für Organisationentwicklung. Dieser Prozess wird auch als digitale Transformation bezeichnet.

Diese, die Hochschule als Organisation verändernde, Entwicklung begründet sich dabei nicht auf die bloße Möglichkeit, neue Technologien für die Organisation zu adaptieren, ist also kein reiner Innovationswille von Innen, auch wenn die Erwartungshaltung der Hochschulangehörigen, insbesondere der Studierenden, an die allgemeine IT-

[1] Christian-Albrechts-Universität zu Kiel, Servicezentrum für Forschung, IT und strategische Innovation, Christian-Albrechts-Platz 4, 24118 Kiel, slorenz@uv.uni-kiel.de
[2] Universität Bielefeld, Dezernat IT/Orga, Universitätsstraße 25, 33615 Bielefeld, frank.klapper@uni-bielefeld.de
[3] ZKI - Zentren für Kommunikation und Informationsverarbeitung e.V., https://www.zki.de/archiv
[4] DINI - Deutsche Initiative für Netzwerkinformation e.V., http://dini.de/veranstaltungen/jahrestagungen/
[5] Mit Service oder Servicebereichen werden all diejenigen Einheiten betitelt, die nicht originär Forschung und Lehre betreiben, insb. Verwaltungseinheiten, Bibliotheken, Rechenzentren oder Multimediazentren.

Infrastruktur und den Digitalisierungsgrad der Hochschule spürbar sind. Entscheidend sind die durch Politik und Forschungsförderer veränderten Rahmenbedingungen für die originären Kernprozesse der Hochschule, die den Veränderungsdruck erhöhen und damit der Notwendigkeit zur digitalen Transformation Nachdruck verleihen [DFG11, DFG12, WRa12]. Der Bolognaprozess etwa führt zu der Notwendigkeit der Einführung von Campus Management Systemen, insbesondere, um den extremen Anstieg an Prüfungen verwalten zu können. Der Kerndatensatz Forschung [WRa13] wird zu der flächendeckenden Einführung von Forschungsinformationssystemen führen, um die Daten für die Forschungsadministration und die Hochschulleitung integriert zugänglich zu machen. Gesetzliche Regelungen zur Einführung der E-Akte führen zur Einführung von Dokumentenmanagementsystemen, usw. Der Ruf nach Steuerung von Forschung und Lehre durch Politik und Forschungsförderer zwingt Hochschulleitungen in die datenbasierte strategische Planung, nicht zuletzt, um im Wettbewerb um die besten Köpfe und Fördertöpfe bestehen zu können.

Hinter jedem der oben genannten Schlagwörter stehen aufwendige Einführungsprojekte, in der Regel mit IT-Bezug und einer hohen organisatorischen Komplexität. Aufwand, Neuartigkeit und Komplexität der Aufgaben im Zusammenhang mit den Themen steigen, was deren Einführung als Projekte rechtfertigt. Insbesondere, wenn die Aufgaben bereichs- und schnittstellenübergreifend angelegt sind, lohnt sich eine Bearbeitung in Projektform. Prägend für viele Projekte sind die Orientierung an durchgängigen Prozessketten, unabhängig von der zugrundeliegenden Linienstruktur, sowie der umfassende Einsatz von IT zur Unterstützung dieser Prozessketten. Aus diesem Kontext heraus sind Unterstützungsleistungen für Projekte in Unternehmen oft am IT- und/oder Organisationsentwicklungsbereich angesiedelt. Viele Hochschulen gehen einen ähnlichen Weg. Die Einführung von Projektstrukturen quer zur Linienorganisation führt zu der Herausbildung von Matrixstrukturen. Die Orientierung an Prozessketten setzt ein Prozessdenken voraus und die Durchführung eines Projekts erfordert Projektmanagementansätze. Matrixstrukturen, Prozessorientierung über Liniengrenzen hinweg und Projektmanagment sind jedoch der klassischen Linienorganisation an Hochschulen fremde Domänen, auch wenn es hier in den letzten Jahren durchaus Bewegung gegeben hat [DeK11]. Die Einführung dieser Domänen in die klassichen Strukturen einer Hochschule erfordert ein umfangreiches Change Managment [EuS05, SSM08]. Unklar bleibt jedoch oft, auf Basis welcher hochschulrelevanten Strategien dieses Change Management durchgeführt wird und mit welchen Methoden die aus dem Change Prozess resultierende Flut gegenwärtiger und zukünftiger Projekte strategisch und operativ sowie personell und finanziell koordiniert werden könnte.

2 Multiprojektmanagement - oder: das Projekt als Daueraufgabe

Um die zukünftige Entwicklung einer Hochschule zu gestalten und zur Anpassung an sich ändernde (interne und externe) Rahmenbedingungen werden, neben den vielen

heute bereits (parallel!) laufenden Projekten, auf Dauer und immer wieder weitere Projekte in größerem Umfang notwendig sein. Das Thema Projektarbeit wird damit zur Daueraufgabe.

Um Projektarbeit als Daueraufgabe nachhaltig in der Organisationsstruktur einer Hochschule zu verankern, bedarf es der Einführung von Methoden zur Bewirtschaftung des Portfolios an laufenden Projekten und vorliegenden Projektvorschlägen, der Entwicklung von Standards und dem Aufbau von Kapazitäten für das Management von Projekten sowie für weitere projektspezifische Unterstützungsleistungen. Organisatorisch empfiehlt es sich, diese Kompetenzen und Ressourcen in einem Projektmanagement-Office (PMO) zu bündeln.

Multiprojekt- oder auch Projektportfoliomanagement bezeichnet die Planung, übergreifende Koordination und Überwachung von mehreren oder allen Projekten einer Organisation. Die Projektkoordination umfasst die Planung, Überwachung und Steuerung und erstreckt sich sowohl auf die „Vorprojektprozesse" der Projektentwicklung und Projektauswahl, auf die Prozesse der eigentlichen Projektabwicklung als auch auf die Nachprojektprozesse der Erfahrungssicherung und des Wissensmanagements [OWL14].

Mit dem Multiprojektmanagement verfolgt eine Organisation das Ziel, *die richtigen Projekte zur rechten Zeit mit den richtigen Ressourcen durchzuführen*, indem zwischen den Sichten der einzelnen Projekte und der globalen Sicht der Organisation und ihrer Gesamtstrategie (sofern vorhanden) vermittelt wird.

Eine wesentliche Aufgabe des Multiprojektmanagements besteht in der Selektion von Projekten nach einheitlichen, nachvollziehbaren Kriterien unter Berücksichtigung der strategischen und organisatorischen Ziele und operativen Notwendigkeiten. Hierbei handelt es sich um eine permanente und zyklisch wiederkehrende Aufgabenstellung, die meist in der Linie, nahe an der Führungsebene, z.B. direkt beim CIO, angesiedelt ist.

Das Mehrprojektmanagement hat insbesondere dafür zu sorgen, dass die aus Sicht der Organisation richtigen Projekte gestartet und erfolgreich durchgeführt werden können. Durch die Auswahl der richtigen Projekte wird die Effektivität des Projektportfolios gesteigert. Effektivität bedeutet, das richtige (hier: Projekt) zu tun. Im Unterschied dazu wird eine hohe Effizienz im Projektportfolio durch die richtige Durchführung der einzelnen Projekte erreicht, also durch Anwendung geeigneter Projektmanagementverfahren auf Einzelprojektebene. Zudem kann ein PMO die neben dem Projektmanagement anfallenden weiteren Aufgaben bündeln, die zusätzliche Fähigkeiten erfordern, welche auf Seiten der Servicebereiche insbesondere bei Projekten mit IT-Bezug typischerweise nicht vorhanden sind (s. Abbildung. 1).

Um in der Universität ein insgesamt stimmiges Gesamtbild (anstelle von mehr oder weniger isolierten Einzellösungen) in den Bereichen Geschäftsarchitektur (Prozesse, Aufbauorganisation), Anwendungsarchitektur (IT-Services, Fachverfahren, Schnittstellen) und Informationsarchitektur (Geschäftsobjekte, Daten) zu erzielen, ist es notwendig, für diese Felder in der Hochschule ein Architekturmanagement aufzubauen

[BIT11]. Es ist sinnvoll, dass das mit diesen Themen betraute Personal in laufenden Projekten bei den Aufgaben Fachkonzepterstellung, Prozessmodellierung, Anforderungsmanagement, IT-Systemauswahl und Umsetzungsspezifikation unterstützt.

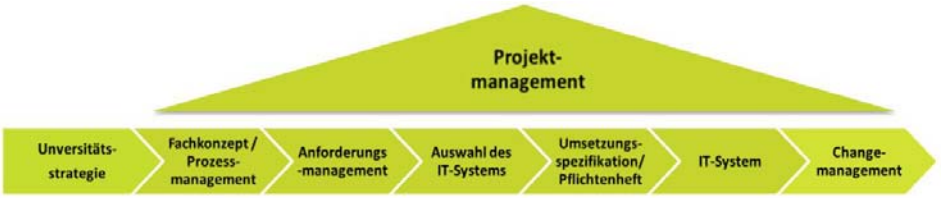


Abbildung 1: Die bei der Realisierung eines Projekts mit IT-Bezug anfallenden projektunterstützenden Aufgaben.

Fast alle Projekte betreffen heute mehrere Organisationseinheiten. Die in Abbildung 1 dargestellten Aufgaben müssen daher organisationseinheiten-übergreifend als Matrixstruktur wahrgenommen werden. Primäres Optimierungsziel sollte das Gesamtergebnis des Projekts und nicht die Optimierung individueller Abläufe zu Lasten der Gesamtlösung sein.

3 Projektmanagement als Service

Für das Projektmanagement und für weitergehende Unterstützung bei der Projektarbeit sollte auf Kompetenzen und Ressourcen zurückgegriffen werden können, die in einem PMO organisatorisch gebündelt sind und die jeweils zu festgelegten Anteilen einzelnen Projekten zugeordnet werden. Operative Aufgaben zur Bewirtschaftung des Projektportfolios (Multiprojektmanagement) werden dem PMO ebenfalls übertragen.

Das Projektmanagementbüro sollte folgendes Leistungsportfolio abdecken:

- Entwicklung von Standards (Methoden, Prozessen, Tools, Templates) für das Projekt- und Mehrprojektmanagement
- Management des Projektportfolios (Mehrprojektmanagement)
- Coaching und Mentoring im Projektmanagement
- Projektcontrolling in Projekten
- Bereitstellung von Projektmanagementressourcen für konkrete Projekte
- Zuarbeit in Projekten (Projektoffice)

Daneben sollte ein enger Bezug zum Architekturmanagement bestehen.

Ein dezentrales Vorhalten von dedizierten Kapazitäten für das Projektmanagement und weitere projektspezifische Sonderaufgaben in den einzelnen Verwaltungseinheiten und weiteren Servicebereichen erscheint wenig sinnvoll, da dann in allen Bereichen entsprechende Kompetenzen aufgebaut werden müssten, diese Kapazitäten jedoch nicht

durchgängig benötigt werden und fast alle Projekte bereichsübergreifend ausgerichtet sind. Die organisatorische Bündelung an einer Stelle dagegen bietet gute Synergiepotentiale, ermöglicht eine hohe Professionalität und kann die bereichsübergreifende Perspektive in die Projektarbeit einbringen (globale Optimierung statt lokaler Optimierung).

4 Das Servicebüro als Organisationsprinzip

Diese Form der bereichsübergreifenden Serviceeinheiten, wie für das PMO beschrieben, ist ein für komplexe Organisationen, wie Hochschulen, wirksames strukturbildendes Prinzip. Die zentrale Bündelung von Kompetenzen, die Entwicklung und Vorhaltung von Standards, die Schulung und Beratung sowie die übergeordnete Steuerung entlang strategischer Vorgaben der Hochschulleitung, wie oben gezeigt am Beispiel des PMO, können helfen, der steigenden Komplexität allgemeiner Anforderungen gerecht zu werden. Zusammengefasst in einem Servicebüro, versetzt sich eine Hochschule am Beispiel des PMO in die Lage, spezifische Strukturen und Serviceangebote aufzubauen, die eine Steuerung in der Fläche überhaupt erst erlauben.

Eine ähnliche Entwicklung der Herausbildung zentraler Strukturen zur professionellen Unterstützung dezentraler Bereiche ist auch bei anderen Themen in Hochschulen zu beobachten, mal als bottom-up Prozess aus Forschung und Lehre heraus, um generalisierbare Strukturen und Services *einmal* zu implementieren und nicht mehrfach (wie E-Learning oder Forschungsdatenmanagement), mal als top-down Prozess, um standardisierte Verfahren und Daten in integrierten Informationssystemen abbilden zu können (wie Campus Management, Forschungsinformation oder Dokumenten-management); die Liste lässt sich beliebig verlängern.

Ähnlich wie im Projektmanagement wurde und wird hier gebündelt Wissen aufgebaut, dass auf einer professionellen Ebene nicht über die gesamte Hochschule verteilt vorgehalten werden kann. Gestartet als Projekt, bilden sich mit der Zeit zentrale Servicebüros oder Kontaktstellen heraus, die den Betrieb entweder selbst übernehmen oder steuern und entsprechend fundiertes Wissen vorhalten, das es früher nicht gab und das früher auch nicht notwendig war. Dieses Wissen kann nun bei Bedarf genutzt, abgerufen oder proaktiv, z.B. als Weiterbildungsangebote, gestreut werden.

Insbesondere Universitäten, die grundsätzlich durch offene Strukturen mit verteilten, zum Teil hochgradig autonom agierenden Einheiten definiert sind, können so den neuen Anforderungen von außen effektiv begegnen. Die Organisation modernisiert und professionalisiert ihr operationales Wissen jenseits von Forschung und Lehre, ohne deren Autonomie zu beschneiden. Darüber hinaus sind Servicebüros personaleffizient, da sie meist nur wenige Vollzeitäquivalente binden.

5 Stabilisierung der Matrixorganisation

Die typische (und logische) Reaktion einer Organisation auf Herausforderungen des Projektmanagements sind Matrixstrukturen, die quer zur Linienorganisation liegen. Nur so können *cross-functional* Teams gebildet werden, die Expertenwissen über die Fachgrenzen hinweg zusammenbringen [DHK96]. Die Gefahr einer Matrixstruktur liegt darin, dass projektbezogene (fachliche) Weisungsbefugnis und linienbezogene dienstrechtliche Weisungsbefugnis kollidieren. Das birgt ein Konfliktpotential für die Organisation, das mit der Anzahl der Projekte und der Belastung der Beschäftigten steigt. Beschäftigte, die sowohl in einer Linienorganisation als auch in Projektstrukturen agieren, müssen u.U. mit mehreren Vorgesetzten klären, welche Aufgaben sie wann übernehmen müssen. Im Idealfall (effizientes Projektmanagement) sind Arbeitsanteile vor Projektbeginn klar definiert. Doch können durch unvorhersehbare Projektverläufe zeitliche Kollisionen entstehen, die zu einem gleichzeitigen Arbeitsaufwand in Projekt und Linie führen. Spätestens dann zeigt sich, dass der Idealfall im Arbeitsalltag oft nicht erreichbar ist. Hinzu kommt das Problem, dass in vielen aktuellen parallel laufenden Hochschulprojekten die Expertise derselben Personen gebraucht wird, eine Person also neben der Linienzugehörigkeit zu mehreren cross-functional Teams gehören kann. Dieses Dilemma ist schwer zu lösen, da Expertenwissen nicht beliebig multiplizierbar ist. Organisationen tendieren im Konfliktfall dazu, wieder zur strikten Linienorganisation zurückzukehren, bis erneut deutlich wird, dass Projekte Matrixstrukturen erfordern. Nicht selten oszillieren Organisationen so über die Jahre zwischen Linien- und Matrixorganisation hin und her und es gibt keine echte Weiterentwicklung (siehe [Obl14, S. 22; Mer12; DHK96]).

Übergeordnete Ansätze, wie ein Multiprojektmanagement und dazu gehörige Strukturen, wie ein PMO, helfen, die Matrixorganisation durch institutionalisierte Querstrukturen soweit zu stabilisieren, das ein Rückfall in die reine Linienorganisation nur schwer möglich ist. Die Transformation der Matrixorganisation in ein drittes, weiterentwickeltes Modell wird selten vollzogen, obwohl Modelle dafür existieren [Obl14].

6 Was kommt nach der Matrix? – Anleihen in der Organisation von Forschung

Um eine echte Weiterentwicklung der Matrixstrukturen in Servicebereichen zu ermöglichen, hilft der Blick in die Organisation von Forschung. Während es in den Verwaltungs- und Infrastruktureinrichtungen einer Hochschule klare hierarchische Strukturen mit Weisungsbefugnis gibt, gibt es diese in der Forschung so nicht. Institute werden in der Regel durch kollegiale Führung der Professorinnen und Professoren geleitet, die geschäftsführende Leitung wechselt turnusmäßig. Forschungsprojekte werden von denjenigen geleitet, die diese Projekte eingeworben haben und es gibt eine Kultur der kollegialen Unterstützung. Diese Führungskultur geht zurück auf die Selbstorganisation von Wissenschaft. Selbstorganisation bedeutet immer auch eine hohe

Eigenverantwortung. In der Forschung beginnt die Eigenverantwortung, wenn die Antragsfähigkeit bei Forschungsförderern gegeben ist, also auf Postdoc-Ebene. Diese liegt deutlich unterhalb der eigentlichen Führungsebene des Instituts. Die verteilte Verantwortung auf mehrere Ebenen ermöglicht eigene Zielsetzungen und eine aktive Beteiligung an übergeordneten Zielen und Strategien im Diskurs und nicht als Vorgabe der *einen* Leitungsebene, s.a. [Bol11] für eine Übersicht zum Modell des *Distributed Leadership*.

Weitere, grundlegende Unterschiede liegen in der Dynamik des Personals und der Dynamik der Themen. Während Serviceeinrichtungen Dauerstellen mit Daueraufgaben vergeben, lebt die Wissenschaft von einem hohen Anteil an jungem und nicht dauerhaft beschäftigtem Personal. Einheiten bauen sich auf und zerfallen u.U. nach einigen Jahren wieder, dieselben Personen organisieren sich in anderen Arbeitsgruppen usw. Der Preis für diese höhere Autonomie und Flexibilität kann daher eine geringere Halbwertzeit von Gruppen oder Themen sein, als das in der klassischen Verwaltung der Fall ist. Neben vielen Nachteilen dieses Modells für das Individuum liegt der Vorteil für die Organisation auf der Hand: Einheiten, die einem regelmäßigen Austausch von Personen und Subthemen unterliegen, können adaptiv auf sich ändernde Rahmenbedingungen reagieren. Serviceeinheiten dagegen mit Daueraufgaben und Dauerpersonal sind zwar zuverlässig in etablierten Prozessen, aber auch träge in der Anpassung an neue Rahmenbedingungen.

7 Die Besetzung der Zwischenräume

In Projekten wie E-Learning, Campus Management, Forschungsinformation oder Forschungsdatenmanagement zeigt sich, wie sehr die ehemals klaren Abgrenzungen zwischen den Serviceeinrichtungen Verwaltung, Bibliothek und Rechenzentrum, aber auch in die wissenschaftlichen Fachbereiche hinein verschwimmen. Zwar bleiben die jeweils klassischen Kernprozesse erhalten, aber es bilden sich Zwischenräume, die mithilfe spezifischer Einheiten, z.B. Servicebüros, gefüllt werden. Diese Einheiten können von allen klassischen Gruppen einer Hochschule (Forschung, Lehre und Verwaltung) genutzt werden und ihnen gleichermaßen nutzen.

Forschungsdatenmanagement (FDM) und E-Learning sind klassische Zwischenraumthemen. Ein Servicebüro FDM baut generalisierbares Wissen über Data Policies, Datenmanagementpläne, gängige FDM-Werkzeuge und Disziplinen unabhängige Infrastrukturen auf und bietet diese als Service in Form von Beratung, Schulung und IT an. Es vernetzt sich national und international und spielt die Informationen zurück in die eigene Organisation. Das Servicebüro FDM bündelt und vernetzt lokale FDM-Aktivitäten in den Fachbereichen und koordiniert die FDM-Angebote und -Beteiligungen der zentralen Einrichtungen wie Rechenzentrum und Bibliothek. Es aggregiert die Anforderungen der Forschungsförderer und berät Forschende bei der Antragsstellung. Die einzelnen Disziplinen können sich auf ihre

fachspezifischen FDM-Probleme konzentrieren und durch eigene Anforderungen das zentrale FDM-Angebot aktiv mitgestalten. Der Aktionsradius des Servicebüros FDM berührt daher Forschungsdaten, wissenschaftliches Arbeiten (Workflows), IT-Infrastruktur, Bibliothekswesen (Metadaten und Repositorien), Forschungsmanagement und -förderung und Projektmanagement und ist damit inhärent transdisziplinär.

Analog verhält es sich mit vielerorts bereits etablierten zentralen E-Learning Services an der Schnittstelle zur Lehre, um bedarfsorientierte Angebote zu generieren, die Lehrende *einfach* nutzen können. Das Serviceportfolio beinhaltet die Bereitstellung einer zentralen Lernplattform und ergänzender Werkzeuge, Schulungen und individuelle Beratung sowohl auf der Ebene der technischen Machbarkeit als auch auf der Ebene des mediendidaktisch sinnvollen Einsatzes. Gemeinsam mit Lehrenden werden die generellen Angebote in die fachspezifische Umsetzung gebracht. Ergänzend kann Medienproduktion angeboten werden. Auch hier gilt die nationale und internationale Vernetzung als Basis der lokalen Innovationsfähigkeit.

Als drittes Beispiel dient der Bereich der Modellierung von Prüfungsordnungen (PO) in Campus Managementsystemen. Organisiert und durchgeführt durch ein zentrales Team wird über die Zeit ein einzigartiges Wissen über die logische Struktur aller PO einer Hochschule aufgebaut. Dieses Wissen kann im Zusammenspiel mit den Fächern, Prüfungsämtern und der zentralen Verwaltung derart genutzt werden, dass PO einem logischen Standard folgen, bereits in ihrer Entstehung auf Modellierbarkeit geprüft werden und Anmeldezeiträume oder Anzahl der möglicher Prüfungswiederholungen so synchronisiert werden, dass Studierende klare und übertragbare Anforderungsstrukturen vorfinden und nicht eine schwer überschaubare Menge von fachspezifischen Einzellösungen. Dieses Wissen kann nur global erzeugt werden, spart in den Fächern Zeit und der Organisation insgesamt Geld und erzeugt zudem Transparenz und Verständlichkeit von PO für Studierende. In dem Team vereint sich IT-Wissen, Prozesswissen, Organisationswissen und juristische Expertise über Prüfungsverwaltung. Das Team arbeitet mit der zentralen Verwaltung genauso wie mit den Fächern und Prüfungsämtern. Es ist interdisziplinäre Schnittstelle und Serviceanbieter.

Alle Beispielen ist gemein, als Projekt zu starten und mit der Zeit nachhaltige (weil nachgefragte) Angeboten und Strukturen zu etablieren. Durch ihr Wirken im Zwischenraum können sie die Hochschulstrukturen im Zusammenspiel mit den jeweiligen dezentralen und zentralen Akteuren nachhaltig verändern und modernisieren.

Interessanterweise werden diese Zwischenräume nicht durch das klassische Verwaltungs- bzw. Servicepersonal gefüllt, sondern durch eine neue Gattung von Wissenschafts- und Hochschulmanagement, die in der Wissenschaft sozialisiert wurden und in den Servicebereich wechseln. Personal mit diesem Profil ist es gewohnt, unter loser Führung projektorientiert und mit hoher Autonomie zu arbeiten und kennt die Bedarfe der Wissenschaft. Dieser Typus kann die Organisation Hochschule in der kommende Dekade entscheidend prägen und durch prozessorientiertes Projekthandeln entlang der strategische Linien die Strukturen der Organisation nachhaltig verändern -

vorausgesetzt, die klassischen Organisationseinheiten bis hinauf zur Hochschulleitung lassen dies zu. Wenn die Organisation als Ganzes es zulässt, dass die Grenzflächen zwischen Forschung, Lehre und Verwaltung aufweichen und Projektarbeit nicht nur in der Forschung als ein Normalfall betrachtet wird, wird sich eine adaptive Struktur etablieren, die flexibel auf Veränderungen reagieren kann.

Multiprojektmanagement kann helfen, diese adaptiven Strukturen zu bilden und zu steuern. Aus Projekten erwachsende Servicebüros agieren an den Grenzflächen zwischen Forschung, Lehre und Service durch Kombination der Vorzüge der Organisationsstrukturen aus Forschung und Lehre, der strukturierten Projektarbeit als Daueraufgabe und Serviceorientierung in cross-functional Teams. In diesem Sinne sind Projekte für die Servicebereiche ein Segen und für die Hochschulen insgesamt eine Chance.

Literaturverzeichnis

[BIT11] BITKOM: Enterprise Architecture Management - neue Disziplin für die ganzheitliche Unternehmensentwicklung, https://www.bitkom.org/files/documents/EAM_Enterprise_Architecture_Management_-_BITKOM_Leitfaden.pdf

[Bol11] Bolden, R.: Distributed Leadership in Organizations: A Review of Theory and Research. International Journal of Management Reviews, 2011, Vol.13(3), S. 251-269, DOI: 10.1111/j.1468-2370.2011.00306.x

[DFG11] DFG – Deutsche Forschungsgemeinschaft: Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme. Empfehlungen der Kommission für IT-Infrastruktur für 2011-2015. http://www.dfg.de/download/pdf/foerderung/programme/wgi/empfehlungen_kfr_2011_2015.pdf

[DFG12] DFG – Deutsche Forschungsgemeinschaft: Die digitale Transformation weiter gestalten – der Beitrag der Deutschen Forschungsgemeinschaft zu einer innovativen Informationsinfrastruktur für die Forschung. Bonn, 2012. http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_digitale_transformation.pdf

[DHK96] Denison, D.R.; Hart, S.L.; Kahn, J.A.: From Chimneys to Cross-Functional Teams: Developing and Validating a Diagnostic Model. The Academy of Management Journal, Vol. 39, No. 4, 1996, S 1005-1023

[DeK11] Degkwitz, A.; Klapper, F. (Herausgeber): Prozessorientierte Hochschule. Allgemeine Aspekte und Praxisbeispiele. Bock + Herchen Verlag, 2011. http://dini.de/fileadmin/docs/Prozessorientierte_Hochschule_2011.pdf

[EuS11] Euler, D.; Seufert, S.: Change Management in der Hochschullehre: Die nachhaltige Implementierung von e-Learning-Innovationen. Zeitschrift für Hochschulentwicklung, 2005.

[Mer12] Mertens, P.: Schwierigkeiten mit IT-Projekten der öffentlichen Verwaltung. Informatik-Spektrum, 35/06, S. 433-446, 2012.

[Obl14] Oblensky, N.: Complex Adaptive Leadership, 2nd Edition, Gower Publishing Company, 2014.

[OWL14] Hochschule Ostwestfalen-Lippe: Handbuch Projektmanagement,https://www.hs-owl.de/fileadmin/aktuelles/rektoratsmitteilungen/Handbuch-Projektmanagement-Web.pdf

[SSM08] Sonntag, K.; Stegmaier, R.; Michel, A.: Change Management an Hochschulen: Konzepte, Tools und Erfahrungen bei der Umsetzung. In: Fisch, R.; Müller, A.; Beck, D. (Herausgeber): Veränderungen in Organisationen. Verlag für Sozialwissenschaften, 2008, S. 415-442. DOI: 10.1007/978-3-531-91166-3_17

[WRa12] Wissenschaftsrat - Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020 (Drs. 2359-12), Juli 2012. http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf

[WRa13] Wissenschaftsrat – Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Januar 2013. http://www.wissenschaftsrat.de/download/archiv/2855-13.pdf

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

tubIT Live – eine komplette Studierendenumgebung "to Go"

Thomas Hildmann[1], Christopher Ritter[2], Nico Suhl[3]

Abstract: Studierende sind heute im Besitz von mehreren mobilen Geräten, die sie mit ins Studium bringen (BYOD). Das Einrichten der für sie geeigneten Lernumgebung stellt sie mit steigender Zahl von Diensten vor eine immer größere Herausforderung. Das hier vorgestellte Linux-System unterstützt das Lernen mit dem PC oder Laptop durch eine automatisierte Personalisierung, die Integration in die Universitätsinfrastruktur, den bereitgestellten Cloudspeicher und ein studienbegleitendes Empfehlungssystem. Das tubIT Live ist sowohl als VM, wie auch in verschiedenen USB-Stickvarianten verfügbar. Ziel ist ferner die Verschmelzung mit der Wartung von PC-Pools.

Keywords: BYOD, Lernumgebung, Infrastruktur, Linux

1 Einleitung

Eine der zentralen Aufgaben der Hochschulen ist die Bereitstellung von Lerninfrastrukturen für Studierende. Waren es vor einigen Jahrzehnten noch vor allem Räumlichkeiten und Literatur, so musste mit zunehmender Digitalisierung und Vernetzung auch entsprechende Computer- und Netzwerkinfrastruktur bereitgestellt werden. Heute ist die zentrale Anforderung der Studierenden nicht mehr die Bereitstellung von Hardware z.B. in Form von PC-Pools, sondern vor allem die Bereitstellung von Bandbreite und Diensten, die zum Lernen benötigt werden. Hardware und ganze digitale Umgebungen, oft bestehend aus diversen Cloud-Diensten werden bereits ins Studium mitgebracht. Mit steigender Zahl der Angebote für Studierende wird auf der anderen Seite das Auffinden der geeigneten Inhalte und Dienste sowie deren Installation, Konfiguration und Verwaltung eine immer zeit- und nervenraubendere Aufgabe für einen Studierenden – ganz besonders zu Studienbeginn. Das hier vorgestellte System bietet daher eine komplette Studierendenumgebung, mit einer Konfiguration durch einmalige Eingabe von Benutzername und Passwort. Es basiert auf einer Linux-Distribution und wird ähnlich dem bekannten Knoppix-Linux als USB-Stick oder wahlweise auch als virtuelle Maschine bereitgestellt. Ziel ist es ein komplett integriertes Ökosystem zu schaffen, wie es Nutzer heute von Anbietern, wie Apple, Google oder Microsoft gewohnt sind und das komplett neben dem bereits etablierten System der Studierenden existieren kann oder aber auch als Ergänzung hierzu genutzt werden kann. Dabei zielen wir bei dieser Lösung nicht auf Smartphones oder Tabletts ab, für die entsprechende Apps bereitgestellt werden. Die Lösung adressiert Laptops, PCs zu Hause oder auch in der Universität, die mit der personalisierten Umgebung genutzt werden können.

[1] Technische Universität Berlin, tubIT, Einsteinufer 17, 10587 Berlin, thomas.hildmann@tu-berlin.de
[2] Technische Universität Berlin, tubIT, Einsteinufer 17, 10587 Berlin, christopher.ritter@tu-berlin.de
[3] Technische Universität Berlin, tubIT, Einsteinufer 17, 10587 Berlin, suhl@tubit.tu-berlin.de

2 Funktionen von tubIT Live

Das System in erster Version wurde mit einem Satz von Funktionen implementiert, die zum einen am häufigsten verwendet werden und zum anderen auch die meisten Anfragen im Support des Rechenzentrums ausmachen. Im Folgenden wollen wir kurz darauf eingehen, wie diese Funktionen jeweils realisiert wurden.

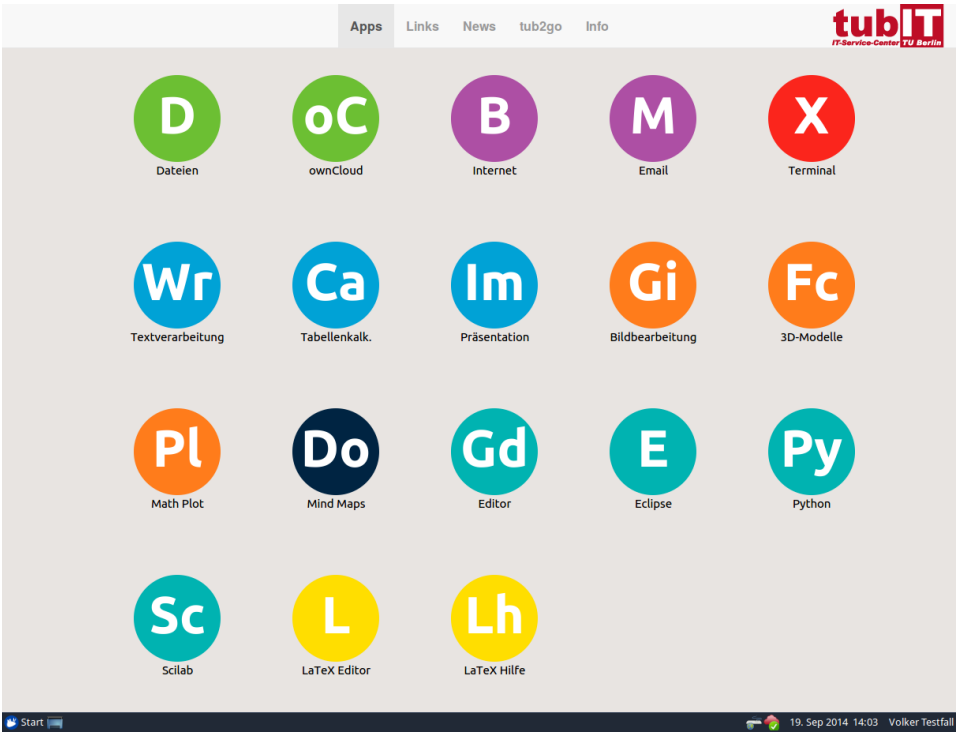


Abb. 1: tubIT Live mit interaktivem Wallpaper

Zentrale Schaltstelle des Systems ist das interaktive Wallpaper, eine HTML-Seite, die den Desktop füllt und Startknöpfe für die wichtigsten Programme enthält sowie weitere über Reiter erreichbare Seiten, die Lesezeichen und ein auf RSS basiertes Mitteilungssystem enthalten.

Die Lesezeichen öffnen jeweils einen Webbrowser und führen auf die zentralen Studierendensysteme für Lehrinhalte, Informationsseiten für Studierende sowie Seiten, die die Funktionalität der mobilen App abbilden (z.B. Indor-Navigation, Mensaplan, Stundenplan).

Zum Erstellen von Texten, Präsentationen und Kalkulationen steht OpenOffice als lizenzkostenfreies System zur Verfügung. Das Grundsystem setzt komplett auf kostenfreie und großteils quellfrei Software, die aber durch kommerzielle Systeme ergänzt werden könnte.

Auch Programmsysteme wie das vollständig vorinstallierte LaTeXwerden von Studierenden verschiedener Studienrichtungen sehr begrüßt. Die Installation einer kompletten LaTeX-Umgebung inklusive geeignetem Editor stellt für einige Studierende offensichtlich noch immer eine nicht zu unterschätzende Hürde dar. Mit dem System ist auch eine offline verfügbare LaTeX-Referenz mit vorinstalliert.

Ein E-Mailprogramm gehört heute wie der Webbrowser zur Standardausstattung. Ebenso wie die Einstiegspunkte für die Webseiten bereits im System enthalten sind, wird das E-Mailkonto bei der Initialisierung des Systems fertig eingerichtet. Mit dem ersten Öffnen des E-Mailprogramms sind alle Einstellungen also bereits gemacht. Weitere bereits vor dem Studium existierende Konten können natürlich hinzugefügt werden, müssen dann jedoch händisch konfiguriert werden.

Der komplette nutzerbezogene Inhalt (Dokumente, Bilder, Multimediadaten) des Systems wird automatisch in den Cloudspeicher (hier ownCloud) synchronisiert. Damit sind die Daten unabhängig vom Speichermedium oder verwendeten Hardwaresystem. Ferner kann auch außerhalb der Plattform über den Cloudspeicher auf die Daten zugegriffen werden.

Bislang war einer der Hauptgründe für die Nutzung von PC-Pools die Bereitstellung von fertig eingerichteten Programmierumgebungen, die für diverse Studiengänge benötigt werden. Viele dieser Programmierumgebungen lassen sich jedoch auch über ein Paket im Linux-System bereitstellen. In einer ersten Version haben wir das System mit gängigen Sprachen und integrierten Entwicklungssystemen ausgestattet. Es ist jedoch vorgesehen, dass abhängig vom Studiengang und ggf. auch besuchter Veranstaltung weitere Pakete installiert werden. Diese können dann auch bereits auf die Anforderungen in der Lehrveranstaltung zugeschnitten sein.

Auch der campusweite Druckdienst steht über die Plattform zur Verfügung und ist vorkonfiguriert. Drucker, die über diesen Dienst angesprochen werden können, sind über den Campus verteilt. So kann sowohl von PC-Pool-Rechnern, als auch vom eigenen Rechner mit tubIT Live aus gedruckt werden.

Eine überschaubare Zahl häufig benötigter Programme wurde bei der Entwicklung des Systems mit Studierenden gemeinsam zusammengestellt und wird mit dem System ausgeliefert. Später ist vorgesehen, in einem vorgegebenen Rahmen eine Auswahl von Programmen oder Alternativen treffen zu können. Grundsätzlich besteht der Mehrwert für die Studierenden jedoch vornehmlich in der redaktionellen Auswahl der Programme, die nicht mehr aus der Vielzahl von Möglichkeiten ausgewählt und evaluiert werden müssen. Auf der anderen Seite bietet die Basis (die Linux-Distribution) für erfahrende Nutzer auch die Möglichkeit weitere Pakete zusätzlich zu installieren. Es ist geplant die Auswahl der vorinstallierten Pakete abhängig vom Studiengang zu variieren.

Eine Liste der im Standard installierten Anwendungen ist im Anhang A aufgeführt.

3 Varianten des Systems

Ein entscheidender Erfolgsfaktor eines Systems, das für die mitgebrachte Hardware der Studierenden (BYOD) gedacht ist, ist die Verfügbarkeit für alle gängigen PC- und Notebook-Modelle und eine möglichst geringe Anforderung an die verfügbaren Ressourcen. Um allen Anforderungen gerecht zu werden, stellen wir tubIT Live in drei Varianten zur Verfügung: zwei USB-Varianten und in einer virtuellen Maschine.

Für die meisten Studierenden ist die Variante USB-Stick die benutzerfreundlichste. Der Stick wird in den Laptop gesteckt, es wird vom Stick gebootet und man kann das komplette System vom Stick benutzen, als wäre es ein eigener Laptop/PC, der nur für das Studium eingerichtet ist. Der Stick kann auch mitgenommen und z.B. im Rechner des/der Kommilitonen/in oder bei den Eltern benutzt werden. Die Bereitstellung muss dabei zwei Hürden überwinden: Die erste ist das Bespielen des USB-Sticks mit dem System und die zweite das Starten von diesem USB-Stick. Für das Bootproblem haben wir entsprechende Anleitungen auf die Webseiten gestellt. Leider gibt es keine wirklich etablierten Standards hierfür.

Um die Herstellung des bootfähigen USB-Sticks jedoch so stark wie möglich zu vereinfachen, haben wir in unserem PC-Pool eine Option hierfür bereitgestellt. Der Nutzer/die Nutzerin wählt aus, dass er/sie sich einen USB-Stick erstellen möchte und wird aufgefordert, diesen in den USB-Port zu stecken. Daraufhin wird das Image auf den Stick kopiert und startet nach einmaliger Initialisierung im eigenen Gerät.

Ferner stehen Anleitungen und Scripte für Windows, OS X und Linux zur Verfügung, um einen Stick zu erstellen. Der kritische Punkt dabei ist, das Image auf das richtige USB-Gerät zu spielen und nicht versehentlich die externe Festplatte zu überspielen.

Leider gibt es mittlerweile auch zwei Firmware-Varianten, die von den Sticks unterstützt werden müssen. Dazu gehört das z.B. von Apple Macs verwendete (U)EFI im Gegensatz zum noch in der PC-Welt verbreiteten „Legacy“ BIOS. Beide Varianten müssen angeboten, beschrieben und gepflegt werden.

Neben den USB-Stick Varianten steht auch eine virtuelle Maschine zur Verfügung, die mit Hilfe des kostenlosen VirtualBox betrieben werden kann. Einige Studierende wollen ihre gewohnte OS-Umgebung nicht verlassen, sondern wünschen sich ein System, dass in einem „Fenster“ oder „Screen“ läuft und auf das bei Bedarf umgeschaltet werden kann. Das private System und das Uni-System sollen nebeneinander laufen. Auch diese Variante wird unterstützt, hat jedoch kleinere Einschränkungen. So kann z.B. die WLAN-Umgebung im virtuellen System natürlich nicht fertig eingerichtet werden, da die Netzverbindung vom Hostsystem stammen muss und an das virtuelle System nur durchgereicht werden kann.

4 Integration in die Lehre

Der integrative Aspekt unserer Lösung geht über die Zusammenstellung häufig genutzter Software und deren Individualisierung durch Konfiguration der Benutzerkonten hinaus.

Um den Studierenden eine gezielte Unterstützung im Rahmen ihres Studiums bieten zu können, muss das Softwarepaket vor allem die für den jeweiligen Studiengang benötigte Software enthalten. Aber auch das alleinige zur Verfügung stellen von Software bietet den Studierenden nur zu einem Teil Unterstützung. Auch aktuelle Informationen bezüglich des Studiums oder einzelner Veranstaltungen werden heutzutage primär in digitaler Form verteilt. Neben den klassischen Mailinglisten werden hier immer häufiger RSS-Feeds eingesetzt. Auch Lerninhalte wie Skripte, Literatursammlungen, Multimedia-Material etc. werden in digitaler Form zur Verfügung gestellt. Die Informationen, welche Inhalte wo zu finden sind, werden häufig dezentral und in unterschiedlicher Form verteilt. Zusätzlich sind viele Inhalte mit einem Passwortschutz versehen, so dass die Studierenden eine Vielzahl an Informationen verwalten müssen um an die jeweiligen Inhalte zu gelangen. tubIT Live wurde daher um ein Modul zur Verwaltung von Inhalten erweitert, das neben den Softwarepaketen auch RSS-Feeds und Lerninhalte besitzt.

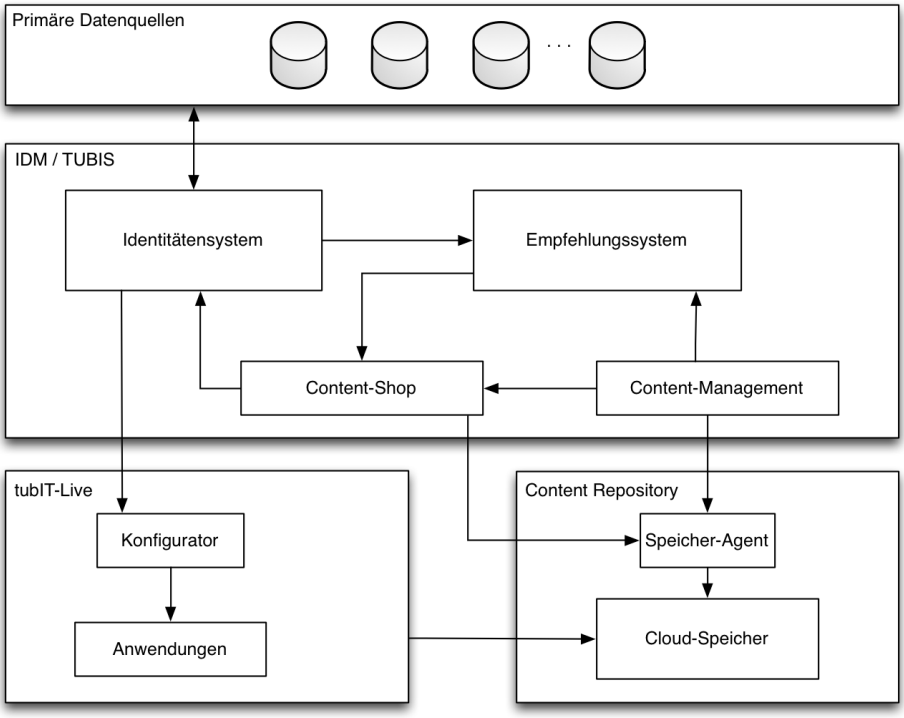


Abb. 2: Zusammenhang: Datenquellen–IDM–tubIT Live

Inhalte, Software und Nachrichtenkanäle können dabei von den Verwaltern der jeweiligen Inhalte mit Metadaten versehen werden, die es dem System ermöglichen diese Komponenten gezielt den gewünschten Empfängern zur Verfügung zu stellen. Mit der direkten Verknüpfung des tubIT Live Systems an das zentrale Identitätsmanagement System existiert zudem die Möglichkeit die Metadaten der Studierenden wie Semesterzahl, Studiengang etc. zu erfassen und Inhalte gezielt zu verteilen. Diese recht statische Zuordnung von Inhal-

ten stellt sich als wenig flexible heraus und bietet den Studierenden keine Möglichkeit auf den Umfang oder die Art der zur Verfügung stehen Inhalte Einfluss zu nehmen. Um dieses Ziel zu erreichen, wurde tubIT Live zusätzlich um ein Empfehlungssystem erweitert. Über ein online Shop ähnliches System können Studierende alle ihnen empfohlenen Inhalte sehen und aus- bzw. wieder abwählen. Darüber hinaus besitzen sie die Möglichkeit auch auf andere Inhalte zu zugreifen, sofern diese nicht auf bestimmte Zielgruppen beschränkt sind. Änderungen in der Auswahl wirken sich dabei im Fall von Inhalten sofort aus oder im Fall von Software beim nächsten Start des Live Systems. In zukünftigen Versionen werden die Empfehlungen nicht mehr ausschließlich auf statischen Werten wie Studiengang, Semester oder ggf. der Auswahl von Veranstaltungen basieren, sondern auf Grundlage einer breiteren Datenbasis getroffen werden. Eine genauerer Beschreibung des Empfehlungssystem wurde in [RBK14] oder mit Bezug auf die Erstellung eines Studienplans in [BRH14] vorgestellt. Mit der Kopplung des tubIT Live an das zentrale Identitätsmanagementsystem war es zudem möglich, die Verteilung des Systems in den Provisioningsprozess der Studierenden zu integrieren, die jede/r Studierende zu Beginn seines Studiums durchlaufen muss. Mit dem Einschreiben bekommt die/der Studierende einen Brief mit einem temporären Passwort zur Erzeugung eines Nutzerkontos. Nach Wahl des Benutzernamens und eines Passworts werden die Ressourcen für den Nutzer freigeschaltet. Hierzu zählen die Netz- und Portalzugänge wie auch Zugang zum ownCloud-Speicher. Am Ende wird unmittelbar auf den Download des USB- bzw. VM-Images verwiesen bzw. auf die Möglichkeit der Erstellung eines USB-Sticks im PC-Pool. Beim ersten Starten des Systems wird nach dem im Provisioningprozess gewählten Benutzernamen und Passwort gefragt. Danach ist das System komplett konfiguriert.

5 Implementierung des Systems

Bei der Implementierung des Systems wurde großen Wert auf die leichte Wart- und Erweiterbarkeit gelegt, so wie auch auf einen hoch verteilten Ansatz. So wurde zu Beginn des Projektes ein GitLab Server [Gi15] aufgesetzt, über den versioniert die gesamte Kontrolle der Konfiguration und der Paketpflege läuft. Ein weiteres Designziel war die Sicherheit des Systems. Es ist selbstverständlich, dass ein USB-Stick entwendet oder verloren werden kann. Befinden sich darauf alle nötigen Zugangsdaten und alle studienrelevanten Dokumente, so ist darauf zu achten, dass ein Zugriff auf diese sensiblen Informationen oder ein Ausspähen der Zugangsdaten nicht möglich ist.

Administriert wird das System über einen Webbrowser. Alle Konfigurationsdateien und Skripte der Bauprozesse werden in git-Repositories verwaltet und dort versioniert. Zur Verwaltung der git-Repositories wird ein GitLab-Server verwendet. Jede Komponente des Systems (Modul) liegt in einem eigenen Repository. Eine Änderung im Repository löst über einen get-Request vom GitLab-Server beim control-Daemon (als Webserver implementiert) ein erneutes Erstellen der Images aus. Der control-Daemon synchronisiert die zum Bau des Livesystems benötigten Module aus den GitLab-Projekten auf den Buildserver und baut neue Images für die unterschiedlichen Varianten des Livesystems mit Hilfe des für den Bauprozess zuständigen Moduls.

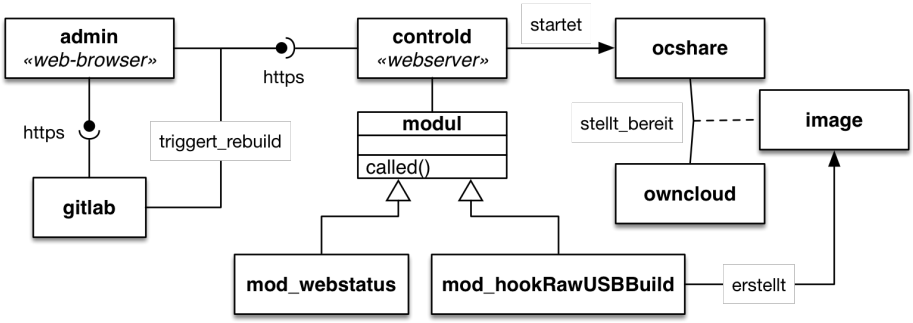


Abb. 3: Klassendiagramm Bauprozess

Die Weboberfläche des control-Daemons erlaubt Administratoren die Übersicht des aktuellen Status, die Ansicht der zuletzt gebauten Images, das manuelle Triggern der Bauprozesse, sowie das Vornehmen von Änderungen an den Einstellungen des Daemons selbst.

Ist ein Image erstellt und getestet, kann es über das ocshare-Script via ownCloud bereitgestellt werden. Somit ist der gesamte Life-Cicle in dem System abgebildet. Änderungen können über die Logs im GitLab nachvollzogen werden. Durch die Versionierung kann jederzeit auf einen älteren Stand zurückgegangen werden, sollten beispielsweise Änderungen den Bauprozess stören. Das gesamte System baut auf dem GitLab-Server und dem build-Server auf, auf dem der control-Daemon läuft. Auf Grund der Modularisierung der verschiedenen Projekte im GitLab ist ein verteiltes Arbeiten in verschiedenen Teams möglich. So kann die Oberfläche z.B. vom Kundendienst gestaltet werden, wohingegen neue Softwarepakete vom UNIX-Team erstellt und hinzugefügt werden.

6 Aufbau des USB-Sticks

Initial, also vor der ersten Nutzung, existiert auf dem USB-Stick eine logische Partition. Auf dieser sind das Tinycoresystem [Ti15], die Konfigurationsskripte, sowie das Basissystem gespeichert. Das Basissystem liegt in komprimierter Form vor und ist nur lesbar.

Zum Start der Systeme wird je nach Variante des Sticks extlinux [ES15] für BIOS-System oder rEFInd [Sm15] für (U)EFI-Systeme als Bootloader genutzt.

Nachdem die Konfiguration des Livesystems über das Konfigurationskript abgeschlossen wurde, befinden sich zwei weitere Partitionen auf dem USB-Stick. Eine FAT32-Partition soll dafür sorgen, dass der USB-Stick nicht versehentlich unter Windows formatiert wird, da die anderen Partition nicht genutzt und geöffnet werden können. Eine entsprechende Textdatei in dieser Partition erklärt den Sachverhalt zusätzlich. Die dritte Partition füllt den kompletten restlichen Speicher des USB-Sticks aus und wird zur Speicherung aller Daten aus dem Livesystem genutzt. Die Partition ist über cryptsetup verschlüsselt, ein darüberliegendes ext4-Dateisystem wird mittels OverlayFS eingehangen, sodass alle am Basissystem vorgenommenen Änderungen in dieser Partition gespeichert werden.

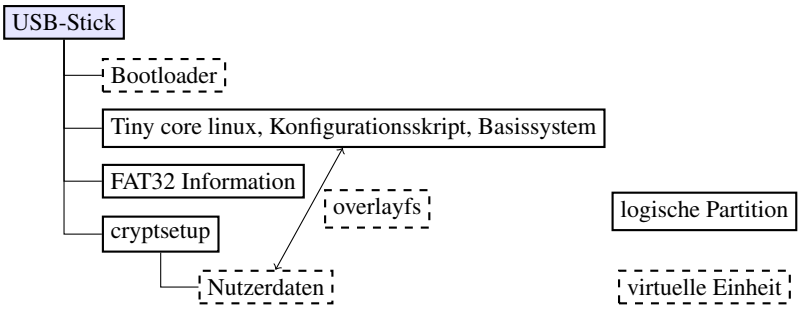


Abb. 4: Aufbau des USB-Sticks

Durch diesen Aufbau wird auch die minimale Größe nutzbarer USB-Sticks festgelegt. Die Größe der initiale Partition, die dem herausgegebenen Image des Livesystems entspricht, gibt die Minimalanforderung an Speicherplatz eines USB-Sticks vor. Der restliche Speicherplatz wird für Änderungen am System und für vom Nutzer erstellte Dateien genutzt. Für das aktuelle Livesystem wird ein USB-Stick mit einer Größe von mindestens acht Gigabyte empfohlen.

Die Variante des Livesystems unter Nutzung einer virtuellen Machine funktioniert analog. Anstelle eines Images für USB-Sticks wird eine virtuelle Festplatte zur Verfügung gestellt, deren Inhalt dem der USB-Images entspricht.

7 Konfigurationsskript

Im Tinycore-Teil des Systems befindet sich als grundlegender Bestandteil das Konfigurationsskript. Tiny Core bietet ein minimales und kleines Linuxsystem (ca. 100MB). Es wird beim ersten Start des Livesystems (Setup) und bei einer Neuinstallation (Recovery) ausgeführt.

Aufgabe des Konfigurationskriptes ist es, das Livesystem vorzubereiten und zu personalisieren. Zunächst werden über eine graphische Benutzeroberfläche die nötigen Benutzerdaten wie Benutzername, Kontoname und Passwort abgefragt.

Zur weiteren Personalisierung wird dann die Partitionierung des Datenträgers erweitert sowie Dateisysteme und deren Verschlüsselung mit dem Benutzerpasswort eingerichtet. Die anschließende Anpassung des Basissystems zur Personalisierung erfolgt über das Overlayfs in den verschlüsselten Bereich. Dabei werden Konfigurationsdateien diverser Programme und Systemtools mit dem Benutzerkonto und -passwort angepasst.

Nach erstmaligem Durchlauf der Konfiguration wird das Startmenü im Bootloader angepasst, sodass das eigentliche Livesystem zum Starten ausgewählt werden kann und das System neugestartet.

Das Konfigurationskript kann jederzeit über den Bootloader ausgeführt werden und das System somit neu aufgesetzt werden.

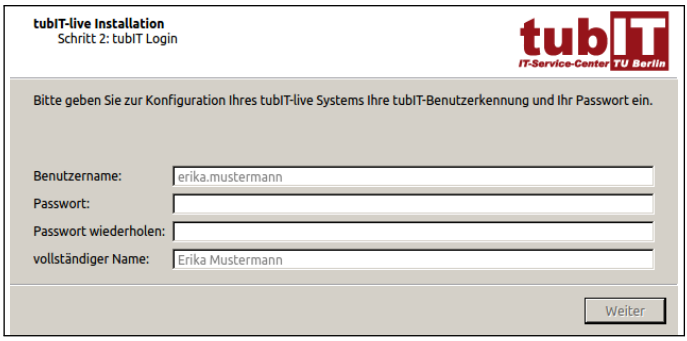


Abb. 5: Graphischer Installationsassistent

8 Cloudspeicher als „Enabling Technology“

Die Kombination des Livesystems mit der Cloudspeicher-Technologie ist von großer Wichtigkeit und macht das System in dieser Form erst so möglich. Der Cloudspeicher erfüllt in diesem Zusammenhang drei wichtige Aufgaben:

1. Unabhängigkeit von der Hardware: Die Studierenden können im Wechsel sowohl mit ihrem Stick, der virtuellen Maschine oder z.B. im PC-Pool arbeiten und haben jeweils Zugriff auf den aktuellen Stand ihrer Dokumente, da diese in die Cloud synchronisiert werden. Ein USB-Stick kann genauso wie jede andere Hardware leicht ersetzt werden. Die Daten werden automatisch auf das neue System synchronisiert.
2. Der systemübergreifende Zugriff auf die Daten bietet eine Schnittstelle zu anderen Plattformen. So steht das Livesystem nicht für sich allein, sondern kann in Kombination mit dem eigenen Windows-System oder einer mobilen Plattform genutzt werden.
3. Das Teilen von Daten über den Cloudspeicher ermöglicht nicht nur eine effiziente Zusammenarbeit mit den Kommilitonen, sondern ist auch eine gute Möglichkeit zur Integration in andere Software. Sie stellt das Recommender-System Lehrinhalte aus einem zentralen Pool über die Teilenfunktion den Studierenden zur Verfügung.

9 Ausblick

Neben der Bereitstellung des tubIT Live Systems wurde an der TU Berlin in die PC-Pools investiert. Ein Linux-basierter Lader kann entweder ein Linux-System lokal starten oder einen Client für die VDI-Umgebung bereitstellen. Weitere Funktionen sind ebenfalls möglich. So ist z.B. das Erstellen der USB-Sticks für das Live-System ebenfalls hierüber realisiert. Ziel ist es nun, das Linux-System der PC-Pools mit dem Live-System zu vereinen, so dass nur noch eine Distribution gewartet werden muss.

Ferner suchen wir andere Hochschulen, mit denen wir das Projekt gemeinsam fortführen können. Aus unserer Sicht bietet es sich an, dass Live-System in Module aufzuteilen,

die an verschiedenen Hochschulen gepflegt werden. Das Live-System lässt sich leicht an die jeweiligen Gegebenheiten an anderen Hochschulen anpassen und kann so als lokales Ökosystem dort bereitgestellt werden.

10 Fazit

Linux-Livesysteme gibt es schon lange. Sie haben jeweils ihre Nischen z.B. als Recovery- oder als Forensikwerkzeug gefunden. Die Verwendung verschiedener mobiler Geräte, wie Smartphones, Tablets und Laptops haben die Nutzer zu einer Multiplattform-Strategie gebracht. Es fällt heute einfacher, mit einem weiteren System zu hantieren, auch wenn es nicht das „Heimatbetriebssystem“ besitzt. Deshalb kommt tubIT Live auch mit einer Oberfläche, die denen der Mobilgeräte nachempfunden ist. Für die Nutzer bietet ein solches Livesystem alles an einem Ort. Für das Rechenzentrum bietet es die Möglichkeit den Support-Aufwand zu senken. Die Universität hat mit einem solchen System die Möglichkeit ein eigenes Ökosystem zu schaffen, wie dies von Apple, Google oder Microsoft bekannt ist mit diversen ineinandergreifenden Diensten, die hier in einer Plattform zusammengefasst werden können, die aber jeweils auch offen für die Nutzung durch andere Plattformen sind. Ohne diese Offenheit würde die Lösung nicht angenommen werden. Ein zentraler Punkt für die Realisierung ist der Einsatz des Cloudspeichers, der die Daten zusammenführt und auf der anderen Seite die Schnittstellen in die anderen Systeme aber auch zu anderen Benutzern bereitstellt.

Literatur

[RBK14] Christopher Ritter, Patrick Bittner, and Odej Kao. Vom BYOE zu GYSE. In Paul Müller, Bernhard Neumair, Helmut Reiser, and Gabi Dreo Rodosek, editors, *7. DFN-Forum - Kommunikationstechnologien, 16.-17. Juni 2014, Fulda, Germany*, volume 231 of *LNI*, pages 67–74. GI, 2014.

[BRH14] Patrick Bittner, Christopher Ritter, and Thomas Hildmann. Get your study plan. In Erhard Plödereder, Lars Grunske, Eric Schneider, and Dominik Ull, editors, *44. Jahrestagung der Gesellschaft für Informatik, Informatik 2014, Big Data - Komplexität meistern, 22.-26. September 2014 in Stuttgart, Deutschland*, volume 232 of *LNI*, pages 1861–1872. GI, 2014.

[Gi15] GitLab Community Edition, GitLab B.V., `https://about.gitlab.com/`, Stand: 19.04.2015

[Ti15] Tiny Core Linux, `http://www.tinycorelinux.net/`, Stand: 19.04.2015

[ES15] Extlinux, Syslinux, `http://www.syslinux.org/wiki/index.php/EXTLINUX`, Stand: 19.04.2015

[Sm15] The rEFInd Boot Manager, Roderick W. Smith, `http://www.rodsbooks.com/refind/`, Stand: 19.04.2015

A Installierte Anwendungen

Die folgende Liste zeigt die im Standard installierten Anwendungen:

Firefox: Safe and easy web browser from Mozilla

Evolution Mail: groupware suite with mail client and organizer

Libre Office: office productivity suite

Gimp: GIMP is an advanced picture editor.

FreeCAD: Extensible Open Source CAx program

KmPlot: mathematical function plotter for KDE

Docear: Docear is a unique solution to academic literature management, i.e. it helps you organizing, creating, and discovering academic literature.

Scilab: Scientific software package for numerical computations

Gummi: GTK+ based LaTeX editor with live preview

Eclipse: Extensible Tool Platform and Java IDE

Pycharm: Python IDE & Django IDE for Web developers

gitg: gitg is a fast git repository browser.

Wer entscheidet? IT-Governance an Hochschulen

Markus von der Heyde[1] und Andreas Breiter[2]

Abstract: Die Steuerung von und Entscheidungen zum Einsatz von Informations- und Kommunikationstechnologien (IuK) sind für Hochschulen zu einer zentralen Aufgabe geworden. Die gesamte Institution umfassende Projekte, Kooperationen zwischen Hochschulen und der stetig wachsende Finanzdruck führten seit mehr als zehn Jahren zu Empfehlungen verschiedener Institutionen: Die Einführung von für Hochschulen geeigneten Steuerungsmodellen z.B. als Chief Information Officer (CIO) war keine Frage, sondern der klare Auftrag.

Im zweiten Jahr der s.g. CIO-Studie wurden Hochschulen mit und ohne CIO-Strukturen gebeten, einen Online-Fragebogen auszufüllen und damit Auskunft über die Steuerung der IT zu geben. Die Schwerpunkte des letzten Jahres wurden dieses Jahr im Wesentlichen beibehalten und nur vom Interview zum elektronischen Fragebogen gewandelt. Zusätzlich haben wir dieses Jahr einen besonderen Fokus auf Details der IT-Governance gelegt: Die jeweils Beteiligten an Entscheidungen zu den einzelnen IT-Domänen sind sowohl für Hochschulen mit CIO-Strukturen als auch für diejenigen ohne untersucht worden. Die Auswertung der insgesamt 39 Datensätze ergibt ein klares Bild, an welchen Stellen die Hochschulen gut oder weniger gut eine IT-Governance umsetzen: Bei der Umsetzung von IT-Rahmenwerken erscheint die Kombination aus Informationssicherheit zusammen mit IT-Prozessmanagement und IT-Servicemanagement als erfolgreich. In Hochschulen mit Integration der Verwaltungsdatenverarbeitung in das zentrale Rechenzentrum sind die IT-Services unklarer und der Wertbeitrag von IT weniger sichtbar.

Keywords: IT-Governance, Hochschule, CIO-Modelle, COBIT.

1 Motivation und Methode

In Unternehmen sind CIO-Modelle lange Jahre erfolgreich etabliert und wissenschaftlich untersucht worden. Die Forderung, ähnliche Modelle in deutschen Hochschulen zu etablieren, ist seit 2001 von verschiedenen Institutionen und Organisationen betont worden [MA01, Ga03, MA06, He08, BK10, GP11 und Hr13]. Hintergrund ist unter anderem die stärkere Fokussierung auf eine Kunden- und Dienstleistungsorientierung der Hochschul-IT und die Einführung von Prozessen entlang von Rahmenmodellen des IT-Service-Management [BF11, Mo05].

Weill und Ross führten 2004 [WR04] die Sichtweise der Entscheidungsdomänen zur Umsetzung einer IT-Governance ein und wiesen nach, dass die Steuerung von IT geeignete Entscheidungsstrukturen benötigt. Das aus diesen Erkenntnissen abgeleitet Framework ist 2008 als ISO Standard [Is08] erschienen. Ebenfalls ist das COBIT 5 PAM

[1] vdH-IT, Rosenweg 7, 99425 Weimar, info@vdh-it.de
[2] Universität Bremen, ifib, Am Fallturm 1, 28359 Bremen, andreas.breiter@ifib.de

ein geeignetes Rahmenwerk, um den Reifegrad der IT-Governance zu bestimmen [GP14].

In europäischen Universitäten wurden erste Analysen, wie weit höhere Bildungseinrichtungen überhaupt diesen Standard implementieren (können), in mehreren Ländern durchgeführt. Zum Beispiel berichten Fernández und Llorens im Rahmen der EUNIS 2009 zu einer landesweit angelegten Initiative, die Steuerungsprozesse an das ISO38500 Framework anzupassen [FL09].

Von Schwabe wurde 2008 ein Vergleich des Reifegrads der Umsetzung von IT-Governance in Österreich, der Schweiz und Deutschland vorgenommen [Sc10] und festgehalten, dass die Umsetzung allenfalls einem wiederholbaren, aber keinem verbesserten Prozess entspricht. In zwei deutschen Hochschulen wurde der Stand 2010 von Börgmann und Bick untersucht und qualitativ beschrieben [BB11].

Quantitative Aussagen zur Umsetzung der IT-Governance wurden erstmals in [He14, WL14] umfassend an deutschen Hochschulen mit einer Interview-Methodik untersucht. In der Nachfolge dieser ersten Studie wurde die Methodik nun in einen elektronischen Web-basierten Fragebogen umgewandelt. Die Ergebnisse sind, da nun eine direkte Angabe der Beurteilung von Fragen durch die Teilnehmer vorgenommen wurde, eigenständig und unabhängig von der Vorgängerstudie zu interpretieren.

Die vorliegende Studie wurde im Zeitraum Mitte März bis Mitte April 2015 durchgeführt. Zur Teilnahme wurden insgesamt 426 Personen von 388 verschiedenen Hochschulen in Deutschland und der Schweiz um die Teilnahme gebeten. Insgesamt 39 Teilnehmende von 37 Hochschulen haben die Fragenbögen vollständig ausgefüllt. Von zwei Hochschulen hat neben dem Leiter des zentralen IT-Servicezentrums (oft auch Rechenzentrum genannt), auch ein CIO bzw. ein Kanzler an der Umfrage teilgenommen. Die Daten wurden auf Plausibilität geprüft, sofern dies möglich und notwendig war korrigiert.

Um den Artikel sprachlich zu vereinfachen, wird oftmals vom CIO und anderen Rollen in männlicher Einzahl gesprochen. Dabei soll nicht angedeutet werden, dass nicht auch weibliche CIOs an der Studie teilgenommen haben und mit den Aussagen gemeint sind. Ebenfalls wird ein Interviewteilnehmer, wenn er oder sie einem CIO Gremium angehört oder dieses repräsentiert, mit CIO bezeichnet.

2 Analyse

Der Fragebogen erfasste insgesamt 254 Einzelvariablen, die von den Teilnehmenden im Durchschnitt in 24 Minuten ausgefüllt wurden. In der anonymen Auswertung einiger Variablen wurden statistische Proben nach Mann-Whitneys U-Test (auch bekannt als Wilcoxon-Rangsummentest) gerechnet. Dabei wurde die Gesamtheit der 39 Teilnehmenden in jeweils zwei Gruppen anhand einer Größe geteilt (z.B. die Drittmittelquote liegt

unter oder 30%). Für die übrigen Daten wurde dann geprüft, ob die Hypothese erfüllt ist, dass die Gruppen sich nicht statistisch unterscheiden. Mit anderen Worten, wenn diese Hypothese nicht erfüllt ist, wird geschlossen, dass die beiden Gruppen nicht einer Grundgesamtheit entstammen, sich also in einer Zielvariablen (z.B. der CIO trägt zu Entscheidungen der IT-Infrastruktur bei) im Rangordnungsverfahren unterscheiden. In diesem Fall kann die Variable (z.B. Drittmittelquote), welche die Gesamtheit in zwei Gruppen teilt, als „Prädiktor" und die Zielvariable (Entscheidungsbeteiligung für Domäne IT-Infrastruktur) als „Indikator" interpretiert werden.

Es werden jeweils die Gruppenstärken, Mittelwerte der zwei Gruppen, der U-Wert des Tests sowie die Standardabweichung der Gesamtstichprobe berichtet. Die statistische Signifikanz des U-Tests wird mit (*) auf dem 5% Niveau der Irrtumswahrscheinlichkeit angezeigt. Mit (**) wird das 1% Niveau ausgezeichnet.

2.1 Charakterisierung der Teilnehmenden

Die Teilnehmenden waren im Schnitt knapp 51 Jahre alt [StdA ~10 Jahre]. Die Verteilung der akademischen Grade ist in Tabelle 1 zu sehen:

	Welche **akademischen Abschluss / Grad** besitzen Sie?							
N=39	Diplom	Dr.	Prof.	Habil.	Master	keinen	Bachelor	Magister
Anzahl Nennungen	26	21	11	3	3	3	1	0
Anteil an Gesamtheit	66,7%	53,8%	28,2%	7,7%	7,7%	7,7%	2,6%	0,0%

Tabelle 1: Anteil der akad. Abschlüsse / Grade der Teilnehmenden. Mehrfachnennungen möglich.

Abgesehen davon, dass Personen mit Professoren- und oder Doktortitel häufiger von größeren Einrichtungen berichteten, waren keine Indikatoren aus der akademischen Karriere abzuleiten. Die aktuellen Rollen der Teilnehmenden zeigt Tabelle 2.

	Welche **aktuellen Rollen** haben Sie an Ihrer Einrichtung inne?									
N=39	RZ-Leitung	CIO	Lehrstuhl-inhaber	Leitung einer anderen zentralen. Einheit	Mitglied im kollektiven CIO	Stabsstelle	IT-Sicherheits-beauftragter	Vizepräsident	Kanzler	Präsident
Anzahl Nennungen	28	18	9	7	5	2	2	1	1	0
Anteil an Gesamtheit	72%	46%	23%	18%	13%	5%	5%	3%	3%	0%

Tabelle 2: Aktuelle Rollen der Teilnehmenden an den Institutionen. Mehrfachnennungen möglich.

2.2 Drittmittelquote

Die Drittmittelquote war in der ersten Studie [He14] aufgefallen, da mit ihr die Entscheidungsbefugnisse des CIOs für IT-Domänen statistisch verbunden schienen. In dieser Studie sollte nach weiteren Anhaltspunkten für die tatsächlichen Zusammenhänge gesucht werden.

Bei 15 Einrichtungen (15/39 = 38,5%) wurden sowohl die Daten zum Gesamthaushalt und als auch zu den darin enthaltenen Drittmitteln angegeben. In diesen Fällen wurde die s.g. Drittmittelquote (DQ) berechnet und zwei Gruppen gebildet (G1 mit DQ<25% [N1=7] und G2 mit DQ>30% [N2=8]). Die Indikatoren aus [He14] sind weiterhin zu erkennen, da die Beteiligung an Entscheidungen in den IT-Domänen auch hier sichtbar blieb [M1=1,3; M2=4,1; StdA=2,3; U=10 von 10 (*)].

Die in dieser Studie neu erfassten Kriterien, wer konkret zu Entscheidungen der IT-Domänen nach Weill und Ross beiträgt, zeigen einen neuen Zusammenhang: Die Gruppe 1 der Hochschulen, die weniger Drittmittel im Vergleich zu ihrem Haushalt ausweisen sagen zu M1=14,3%, dass der CIO zu Entscheidungen der IT-Infrastruktur beiträgt. Die andere Gruppe 2 bescheinigt dem CIO hier Beteiligung an Entscheidungen [M2=87,5%; StdA=49,9%; U=7,5 von 10 (*)]. Nicht unerwartet besteht die Gruppe 2 (mit hoher DQ) vollständig aus Universitäten. In der Gruppe 1 sind nur M1=29% der Hochschulen Universitäten [StdA=47%; U=8 von 10 (*)]. Daher ist es ebenfalls zu erwarten, dass Gruppe 2 deutlich mehr Studierende ausweisen kann [M1=6297; M2=18803; StdA=10625; U=7,5 von 10 (*)]. Abschließend ist festzuhalten, dass die beiden Gruppen sich im Vorhandensein einer eigenständigen Bibliotheks-IT unterscheiden [M1=14,3%; M2=87,5%; StdA=49,4%; U=7,5 von 10 (*)].

Die Vorgängerstudie aufgreifend wurde untersucht, ob die Drittmittelquote abermals als Indikator statistisch signifikant erscheint. Wir unterteilten die Hochschulen, die Zahlen für die Berechnung der DQ angegeben haben, danach, ob der CIO an den Entscheidungen in den IT-Domänen beteiligt ist: G1 mit keiner Domäne [N1=6]; G2 mit mindestens der Beteiligung an der Entscheidung einer Domäne [N2=9]. Die Gruppen unterscheiden sich nicht nur in der Drittmittelquote [M1=14,8%; M2=35,5%; StdA=15,0%; U=5 von 10 (*)] sondern auch in der Sichtbarkeit der Sachmittel und Personalstellen für IT in der Verwaltung.

Wenn die Institutionen danach unterschieden werden, ob eine eigenständige HPC Abteilung existiert, ergibt sich ein weiterer Bezug. Die Gruppe 1 [N1=10] ohne HPC unterscheidet sich von der Gruppe 2 [N2=5] mit HPC in Bezug auf die Drittmittelquote [M1=21,2%; M2=39,3%; StdA=15,0%, U=7 von 8 (*)]. Die um rund 17% höhere DQ ist sicherlich auch durch die Ausgaben im HPC Bereich selbst zu begründen.

Der unerwartete Faktor des stärkeren CIO Beitrags zu Infrastrukturentscheidungen könnte in Zusammenhang mit den Entscheidungen zu HPC Infrastrukturen stehen und damit beide Beobachtungen erklären.

2.3 Wirkungen und Umsetzung der IT-Governance Rahmenwerke

Zwei Fragen zielten auf die Wirksamkeit von IT-Governance Rahmenwerken (z.B. Cobit, COSO, ISO38500) an den befragten Hochschulen. Teilt man die Gesamtheit in die Gruppen danach ein, ob die Teilnehmer selbst sich in ihrer Arbeit an diesen Rahmenwerken orientieren [N1=32 (nein); N2=7 (ja)], so ergeben sich nur sehr leichte Effekte, die gerade die Signifikanzschwelle erreichen. Zum einen beteiligt sich der CIO in Gruppe 2 häufiger an den Entscheidungen zur IT-Strategie [M1=37,5%; M2=85,7%; StdA=49,9%; Z=58 von 58 (*)]. Zum anderen tragen in Gruppe 2 die Fakultäten mit Informationen zu den Entscheidungen in der IT-Infrastruktur bei [M1=37,5%; M2=85,7%; StdA=49,9%; Z=58 von 58 (*)]. Obwohl die Statistik zu beiden Indikatoren identisch ausfällt, tragen jeweils andere Hochschulen dazu bei.

Leider ergab die zweite Frage, ob IT-Governance Rahmenwerke übergreifend zum Einsatz in den Hochschulen kommen, keine Ergebnisse. Nur ein Teilnehmer berichtete von seiner Institution, dass diese sich nach IT-Governance Rahmenwerken ausrichte. Ebenfalls ernüchternd berichteten nur zwei Hochschulen, dass letztlich der CIO übergreifende IT-Entscheidung treffe. Dies kann sowohl auf fehlende Sensibilität für das Thema hinweisen als auch auf die schlechte Passung von Rahmenwerken auf die Hochschulstrukturen.

Immerhin dokumentieren die Hochschulen, die angeben, einen CIO zu haben [N1=24] in Gegensatz zu den anderen [N=15] eine statistisch signifikante Beteiligung des CIOs an Entscheidungen in **allen** IT-Domänen nach Weill und Ross.

HS **mit** CIO	Beteiligung an Entscheidungen nach IT-Domänen									
N=24	Präsidium	Vizepräsident	Kanzler	CIO	RZ-Leitung	Sonstige Abt.-Leiter	Fakultäten	Mitglieder des höchsten IT Gremiums	Jeder	Dies ist nicht geregelt
IT-Strategie	71%	13%	33%	75%	38%	4%	13%	21%	0%	4%
IT-Architektur	25%	0%	8%	79%	75%	4%	13%	13%	0%	8%
IT-Infrastruktur	46%	0%	29%	71%	88%	13%	17%	17%	0%	4%
IT-Applikationen	50%	21%	54%	83%	79%	42%	29%	21%	4%	13%
Priorisierung und Investition	63%	8%	54%	63%	75%	13%	13%	13%	0%	4%

Tabelle 3: Übersicht der Beteiligung an Entscheidungen in den IT-Domänen nach Weill und Ross an 24 Hochschulen, die **eine** CIO-Form etabliert haben.

Die jeweils zu den IT-Domänen maßgeblich beitragenden Personengruppen sind in der Übersicht der Entscheidungsträger an Hochschulen mit CIO in Tabelle 3 und in der

folgenden Tabelle 4 grün unterlegt. Zeilenweise sind die jeweils am häufigsten Beteiligten markiert.

HS **ohne** CIO	Beteiligung an Entscheidungen nach IT-Domänen									
N=15	Präsidium	Vizepräsident	Kanzler	CIO	RZ-Leitung	Sonstige Abt.-Leiter	Fakultäten	Mitglieder des höchsten IT Gremiums	Jeder	Dies ist nicht geregelt
IT-Strategie	40%	0%	33%	0%	60%	0%	7%	13%	0%	27%
IT-Architektur	13%	7%	27%	0%	87%	7%	7%	7%	7%	20%
IT-Infrastruktur	13%	0%	27%	0%	73%	13%	13%	13%	7%	13%
IT-Applikationen	33%	13%	53%	0%	67%	27%	13%	13%	7%	20%
Priorisierung und Investition	40%	20%	67%	0%	60%	13%	13%	0%	13%	13%

Tabelle 4: Übersicht der Beteiligung an Entscheidungen in den IT-Domänen nach Weill und Ross an 15 Hochschulen, die **keinen** CIO haben

Im auffallenden Unterschied dazu sind in Tabelle 4 bei Hochschulen ohne CIO die fehlende Regelungen sichtbar bzw. die „Anarchie“ (letzten beiden Spalten), dass jeder an Entscheidungen mitwirkt. Aus den Markierungen, in welchen Domänen welche Personengruppen am stärksten an den Entscheidungen beteiligt sind, ist ersichtlich, dass bei der IT-Strategie die Entscheidungen, wenn kein CIO an der Hochschule berufen wurde, vom Rechenzentrum in Zusammenarbeit mit dem Präsidium gefällt werden. Für IT-Infrastruktur und IT-Architektur wird der RZ-Leiter ggf. vom CIO unterstützt und fällt die Entscheidungen mit ihm zusammen. Wiederum bei Hochschulen ohne CIO übernimmt der Kanzler die Entscheidungsbefugnisse statt eines CIOs für die IT-Applikationen. Für die Priorisierung werden Entscheidungsbefugnisse von Präsidium und Kanzler an einen CIO übertragen, wenn es diesen gibt. Das Rechenzentrum bleibt für alle Domänen (mit Abschwächung für die IT-Strategie, wenn es einen CIO gibt) an den Entscheidungen in allen Konstellationen beteiligt.

Bemerkenswert ist zusätzlich, dass bei CIO-Modellen in Form eines Gremiums [N1=5] häufiger als bei anderen Modellen [N2=17] die Entscheidungen zur IT-Architektur beim Präsidium verbleiben [M1=80%; M2=12%; StdA=45%; U=13,5 von 17 (*)]. Die Entscheidung zu IT-Applikationen sind bei CIO-Gremien sogar vollständig beim Präsidium [M1=100%; M2=35%; StdA=50%; U=15 von 17 (*)].

2.4 Vergleich von Hochschulen mit „not for profit“-Organisationen

„Not for profit“-Organisationen sind, wie Hochschulen auch, nicht auf die Maximierung

von Gewinn ausgerichtet. Daher werden für die Steuerung der IT oftmals andere Indikatoren und Zielgrößen genutzt.

Vergleicht man die Gesamtheit der Hochschulen mit den „not-for-profit"-Organisationen (kurz nfpO) aus [WR04, Figure 7-5, P.202], so sind einige Unterschiede sichtbar. In Tabelle 5 sind die primär für Entscheidungen gefundenen Modelle blau für den Input und grün für Entscheidungen hinterlegt.

Für den Beitrag von Informationen für die Entscheidungen unterscheiden sich die hier untersuchten Hochschulen in der IT-Strategie, IT-Infrastruktur, den Geschäftsapplikationen und der Priorisierung/Investition nach dem gleichen Muster von den nfpO: Der bei den nfpO vorherrschende federale Beitrag ist bei den Hochschulen zu annähernd gleichen Teilen federal und als IT-Duopoly ausgeprägt. Bei dem Informationsbeitrag für die IT-Architektur ist das Bild genau umgekehrt. Hochschulen haben hier deutlich häufiger das federale Modell gewählt und die nfpO teilen sich zwischen IT-Duopoly und federalem Ansatz auf.

		IT-Strategie		IT-Architektur		IT-Infrastruktur		Geschäfts-appl.		Invest. und Prio.	
		Input	Ents.	Input	Ents.	Input	Ents.	Input	Ents.	Input	Ents.
Business Monarchy	HS	0%	18%	0%	8%	0%	10%	0%	8%	3%	18%
	nfpO	0%	35%	0%	5%	0%	11%	0%	15%	0%	36%
IT Monarchy	HS	10%	8%	13%	49%	15%	28%	5%	8%	10%	3%
	nfpO	0%	9%	9%	67%	4%	58%	0%	3%	0%	5%
Feudal	HS	0%	3%	0%	3%	0%	3%	3%	3%	0%	3%
	nfpO	1%	1%	0%	0%	0%	1%	3%	20%	0%	1%
Federal	HS	41%	15%	28%	8%	31%	21%	41%	38%	31%	26%
	nfpO	86%	15%	49%	5%	66%	8%	82%	43%	95%	24%
IT Duopoly	HS	36%	49%	44%	31%	38%	36%	31%	38%	46%	46%
	nfpO	12%	37%	42%	23%	30%	22%	14%	15%	5%	31%
Anarchy	HS	13%	8%	15%	3%	15%	3%	21%	5%	10%	5%
	nfpO	0%	0%	0%	0%	0%	0%	1%	4%	0%	3%

Tabelle 5: Archetypen nach [WR04] für Entscheidungen und Beitrag durch Informationen zu diesen. Die Zahlen der Hochschulen befinden sich oben links und die Zahlen aus [WR04] mit nfpO bezeichnet unten rechts. Alle Spalten summieren sich innerhalb der Gruppen zu 100%.

Die Situation bei den Archetypen für die Entscheidungen innerhalb der IT-Domänen ist wesentlich heterogener. Hier sind nur für die IT-Architektur keine wesentlichen Unterschiede zwischen Hochschulen und den nfpO zu sehen. Für IT-Strategie, IT-Infrastruktur und die Entscheidungen zu Priorisierung und Investitionen konzentrieren sich die Hochschulen wesentlich häufiger auf IT-Duopoly als die nfpO. Für die

Entscheidungen zu Geschäftsapplikationen ist das Bild identisch zu dem Beitrag als Information.

Weiterhin ist auffallend, dass bei den Hochschulen die Reife der IT-Governance nicht so hoch scheint, wie bei den von Ross und Weill untersuchten nfpO, da sowohl für den Informationsbeitrag als auch für die eigentlichen Entscheidungen in allen IT-Domänen der Anteil der Anarchie höher ausfällt.

2.5 Zusammenhänge zwischen IT-Governance, IT-Sicherheit und IT-Servicemanagement

Die Nutzung von Rahmenwerken scheint offensichtlich und kann durch diese Studie bestätigt werden.

	Folgende Frameworks werden in meiner Institution für übergreifende Aufgaben angewendet:		An folgenden Frameworks orientiert sich meine eigene Arbeit:	
N=39	absolut	Anteil	absolut	Anteil
IT-Governance (Cobit, COSO, ISO38500, …)	1	3%	7	18%
IT-Architektur (TOGAF, SOA, …)	4	10%	3	8%
Informationssicherheit (BSI, IT-Grundschutz, ISO2700x, …)	19	49%	20	51%
IT-Servicemanagement (ITIL, FitSM, MOF, ISO20000, …)	14	36%	21	54%
Projekte (PRINCE2, PMP, IPMA, PMBOK, ICB, HERMES, ISO21500, …)	11	28%	11	28%
Prozesse (GPM, BSM, ….)	11	28%	12	31%
Qualität (EFQM, Six Sigma, LQM, EN ISO 9001, ...)	5	13%	4	10%
Entwicklung (V-Modell, Agile, Scrum, …)	9	23%	7	18%
Projekt-Portfolio-Management / Innovations-Management	8	21%	11	28%
Organisationsbezogenes Change-Management	18	46%	16	41%

Tabelle 6: Häufigkeit der Verwendung von Rahmenwerken in absoluten Zahlen und Anteilen.

Diejenigen Hochschulen, die eines der IT-Sicherheitsrahmenwerke institutionsweit einsetzen [N1=19] unterscheiden sich von denjenigen, die dies (noch) nicht einsetzen [N2=20], in mehreren Faktoren. Gleichzeitig zur IT-Sicherheit kommt das IT-Servicemanagement [M1=57,9%; M2=15,0%; Var=48,0%; U=108,5 von 119(*)] und ein Prozess-Framework [M1=47,4%; M2=10,0%; Var=45,0%; U=119 von 119 (*)] zum Einsatz. Insbesondere das IT-Servicemanagement scheint eine starke Orientierungshilfe für Entscheider zu sein. Diese beiden Ergebnisse werden in der umgekehrten Richtung ebenfalls statistisch signifikant bestätigt.

2.6 Verwaltungsdatenverarbeitung als Teil des Rechenzentrums

Die seit Jahren zunehmende Integration von IT Abteilungen der Verwaltung in das zentrale Rechenzentrum ist deutlich zu beobachten. Von den 39 Teilnehmern haben nur noch 12 Hochschulen [N1=12] ein eigenständige Verwaltungsdatenverarbeitung. Bei den anderen [N2=27] sind die Bereiche heute verschmolzen. Die zwei Gruppen unterscheiden sich insbesondere in der Bewertung der Services ihrer Hochschulen:

	Ihnen sind die IT-Sachmittel bekannt…					
	In Ihrem eigenen direkten Verantwortungs-bereich	In der Verwaltung	In einzelnen Fakultäten / Fachbereichen	In allen Fakultäten / Fachbereichen	Für alle Drittmittelprojekte	Der gesamten Institution
keine Beteiligung des CIO an Entscheidungen in einer IT-Domäne [N1=18]	78%	28%	6%	0%	0%	17%
Beteiligung des CIO an Entscheidungen in mindestens einer IT-Domäne [N2=21]	95%	86%	19%	19%	10%	19%
StdA	33%	49%	33%	30%	22%	38%

Tabelle 7: Sichtbarkeit der Ausgaben für IT-Sachmittel in Abhängigkeit, ob der CIO an Entscheidungen in den IT-Domänen beteiligt ist

Die erste Gruppe (Hochschulen mit unabhängiger Verwaltung-DV) ist auf einer Skala von 1=„stimme vollständig zu“ bis 5=„stimme gar nicht zu“ eher unschlüssig, ob IT Service mit Kunden in SLAs abgestimmt werden [M1=2,8]. Die Hochschulen mit

integrierter Verwaltungs-DV sind dagegen eher davon überzeugt, dass die Services nicht mit den Kunden abgestimmt werden [M2=4,2; StdA=1,1; U=52,5 von 78 (**)]. Die Situation ist sehr ähnlich für die Frage, ob die Kosten der IT-Services ermittelt und kommuniziert werden [M1=3,25; M2=4,22; StdA=1,07; U=69 von 78 (**)]. Auch für die Frage, ob die IT-Services sich aus dem aktuellen und zukünftigen Mehrwert für die Kernprozesse ableiten, zeigt die Gruppe 1 eine positivere Haltung [M1=2,42; M2=3,15; StdA=1,00; U=90,5 von 97 (*)]. Auch zur Gegenfrage, ob IT-Services einen unklaren Wertbeitrag haben, ist die Situation stimmig [M1=3,58; M2=2,41; StdA=1,12; U=71 von 78 (**)]. Die Hochschulen mit integrierter Verwaltungs-DV sehen hier eine Zustimmung, d.h. es besteht kein klarer Wertbeitrag der IT-Services.

Wir möchten die Anzahl der IT-Domänen, bei denen der CIO an der Entscheidung beteiligt ist, als Prädiktor an dieser Stelle aufgreifen. Wir unterteilen die Gesamtheit in die Gruppe der Hochschulen, die keinen CIO hat bzw. dem CIO keine Befugnisse einräumt [N1=18] und diejenigen, die dem CIO mindestens für eine IT-Domäne eine Beteiligung an der Entscheidung einräumen [N2=21]. Die Sichtbarkeit der Sachmittel, die für IT Aufgaben ausgegeben werden, ist in allen Bereichen (siehe Tabelle 7) absolut erhöht, im Bereich der Verwaltung sogar statistisch hoch signifikant [M1=27,8%; M2=85,7%; StdA=49,2%; U=98 von 119 (**)].

Die Zahlen, welche offensichtlich einigen CIOs bereits vorliegen, sind auch für weitere Initiativen wie zum Beispiel einem IT-Benchmarking vom hohen Wert. Die Situation in der Sichtbarkeit für IT Personal ist fast analog. Im Bereich der Drittmittel ist keinerlei Sichtbarkeit gegeben, dafür ist die Sichtbarkeit in den Fachbereichen leicht erhöht. Statistisch ist auch hier der Unterschied im Bereich der Verwaltung signifikant.

3 Zusammenfassung

Die erste Auswertung der zahlreichen Indikatoren ergab, dass in großen Teilen die Ergebnisse (z.B. Drittmittelquote als Prädiktor und Indikator) aus der ersten Studie bestätigt werden konnten. Zusätzlich konnten wir die Archetypen in dem Beitrag für die Entscheidungen in den jeweiligen IT-Domänen ermitteln. Im Vergleich zu den bei „not for profit“-Organisationen gefundenen Archetypen sind Hochschulen in fast allen IT-Domänen eher als IT-Duopoly als federal aufgestellt.

Der Nachweis von statistischen Zusammenhängen für Hochschulen, die IT-Governance Rahmenwerke eingeführt haben oder in denen der Teilnehmer sich selbst an solchen Rahmenwerken orientiert, konnten nicht erbracht werden, da zu wenige Hochschulen angaben, mit entsprechenden Rahmenwerken zu arbeiten. An den Institutionen, an denen ein Rahmenwerk zur Informationssicherheit etabliert ist, kommen signifikant häufiger IT-Servicemanagement und IT-Prozess-Management zum Einsatz. Entsprechende Rahmenwerke eignen sich offensichtlich gut zum gemeinsamen Einsatz.

Besonderes Augenmerk haben wir in dieser Analyse auf die Interaktion zwischen

Entscheidungsstrukturen und der Integration von Verwaltungs-DV mit dem Rechenzentrum gelegt. Hier überraschen die Hochschulen, die eine Integration bereits vollzogen haben, damit, dass IT-Services in ihrer Formulierung von Kosten und Nutzen unklarer erscheinen. Dies kann als Indikator dafür gesehen werden, dass bei der Integration der Verwaltungs-DV in klassische Rechenzentren das Profil der Gesamteinrichtungen noch nicht ausreichend klar formuliert wurde.

Wenn eine Hochschule ein CIO-Modell etabliert, erscheint es nach den hier präsentierten Daten ratsam, die Person – ggf. auch ein Gremium – an den Entscheidungen zu mindestens einer der IT-Domänen zu beteiligen.

Danksagung

Wir danken den Teilnehmern der Studie für das Vertrauen und die Unterstützung bei der Studie. Wir danken besonders Frau Michaela Meyer von der Universität Bremen für die Umsetzung des Online-Fragebogens.

Literaturverzeichnis

[BB11] Katrin Börgmann, Markus Bick: IT-Governance in deutschen Hochschulen - eine qualitative Untersuchung. Hochschulmanagement (HM) 6(2):47–53, 2011, http://www.universitaetsverlagwebler.de/inhalte/hm-2-2011.pdf, Abruf zuletzt am 30.04.2015

[BF11] Andreas Breiter, Arne Fischer: Implementierung von IT Service Management in der öffentlichen Verwaltung. Nationale und internationale Fallstudien und deren vergleichende Analyse. Berlin: Springer, 2011.

[BK10] Werner Bröcker, Frank König: Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme, Empfehlungen der Kommission für IT-Infrastruktur für 2011–2015, Deutsche Forschungsgemeinschaft 2010, http://www.dfg.de/download/pdf/foerderung/programme/wgi/empfehlungen_kfr_2011_2015.pdf, Abruf zuletzt am 30.04.2015

[FL09] Antonio Martínez Fernández, Faraón Largo Llorens: "An IT-Governance framework for universities in Spain". In: Proceedings of the EUNIS 2009 Conference: June 23rd to 26th 2009, University of Santiago de Compostela, 2009. ISBN 978-84-9887-138-8, http://www.gti4u.es/pdf/an_it_governance_framework_for_universities_in_spain.pdf, Abruf zuletzt am 30.04.2015

[Ga03] Rudolph Gardill: Empfehlungen des ZKI e.V. zur Etablierung eines CIO-Funktionsträgers in Hochschulen, Vorlage für die 13. Mitgliederversammlung 2004 in Würzburg, 2003, https://www.zki.de/fileadmin/zki/Gremien/vorstand/veroeffentlichungen/CIO-Papier.pdf, Abruf nur für Berechtigte.

[GP14] Tobias Grönert, Jens Pöppelbuß, Andreas Breiter (2014): Reifegradbestimmung der IT-Governance: Eine Fallstudie zur Anwendbarkeit des COBIT 5 PAM in der öffentlichen Verwaltung. in: E. Plödereder, L. Grunske, E. Schneider & D. Ull (Hrsg.), Informatik 2014. Big data - Komplexität meistern, Stuttgart: Springer, S. 1513-1525.

[GP11] Simone Görl, Johanna Puhl, Manfred Thaller: Empfehlungen für die weitere Entwicklung der Wissenschaftlichen Informationsversorgung des Landes NRW, ISBN: 978-3-8442-0694-4, Köln, 2011

[He08] Markus von der Heyde et al.: IuK-Strukturkonzepte für Hochschulen - Empfehlungen zu Zielsetzung, ZKI e.V., Aufbau und Inhalt, http://www.zki.de/fileadmin/zki/Publikationen/IuK-SK_Gesamt-V1_0_02.pdf, Abruf zuletzt am 30.04.2015

[He14] Markus von der Heyde: CIO-Modelle an deutschen Hochschulen Netzwerk der entscheidenden Faktoren, LNI 232, Seiten 1815-1823, 2014/9, http://cs.emis.de/LNI/Proceedings/Proceedings232/1815.pdf, Abruf zuletzt am 30.04.2015

[Hr13] Entschließung der 13. Mitgliederversammlung der HRK am 20. November 2012 in Göttingen: Hochschule im digitalen Zeitalter: Informationskompetenz neu begreifen - Prozesse anders steuern, erschienen in Beiträge zur Hochschulpolitik 1/2013, http://www.hrk.de/fileadmin/redaktion/hrk/02-Dokumente/02-10-Publikationsdatenbank/Beitr-2013-01_Informationskompetenz.pdf, Abruf zuletzt am 30.04.2015

[Is08] ISO/IEC 38500:2008 Corporate Governance of Information Technology", ISO/IEC. 2008, from: www.iso.org

[MA01] Joachim Mönkediek, Peter Arends: Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme, Empfehlungen der Kommission für Rechenanlagen für 2001–2005, Deutsche Forschungsgemeinschaft 2001, http://www.dfg.de/download/pdf/foerderung/programme/wgi/wgi_kfr_empf.pdf, Abruf zuletzt am 30.04.2015

[MA06] Joachim Mönkediek, Peter Arends: Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme, Empfehlungen der Kommission für Rechenanlagen für 2006–2010, Deutsche Forschungsgemeinschaft 2006

[Mo05] Horst Moog: IT-Dienste an Universitäten und Fachhochschulen. Reorganisation und Ressourcenplanung der der hochschulweiten IT-Versorgung. Hannover: HIS. Hochschul-Informations-System GmbH, 2005.

[Sc10] Gerhard Schwabe: IT-Governance an Universitäten in Deutschland, Schweiz und Österreich, in ZKI-Herbsttagung 2010 – IT-Governance und IT-Infrastruktur-management in europäischer Ausprägung, September 2010, http://www-docs.tu-cottbus.de/ikmz/public/files/Veranstaltungen/Tagungswoche_09_2010/Vortraege/ZKI/02_Schwabe_IT-Governance_VM_Teil_2.pdf, Abruf zuletzt am 30.04.2015

[WL14] Martin Wimmer, Ulrich Lang [Publisher]: „CIOs und IT-Governance an deutschen Hochschulen", ZKI e.V., http://www.zki.de/fileadmin/zki/Publikationen/ZKI_CIO-Studie_final.pdf, Abruf zuletzt am 30.04.2015

[WR04] Peter Weill, Jeanne W. Ross: IT Governance: How Top Performers Manage IT Decision Rights for Superior Results". Harvard Business School Press, Boston 2004, ISBN 978-1591392538.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

IT-Dienste im Kontext von Hochschulprozessen – Eine IT-Landkarte zur Unterstützung des IT Servicemanagements

Christian Erfurth[1], Ivonne Erfurth[2]

Abstract: Deutsche Hochschulrechenzentren beschäftigen sich derzeit immer mehr mit der Umsetzung von IT-Servicemanagement (ITSM). Erste Schritte sind die Etablierung einzelner ITSM-Prozesse, wie Incident Management und damit verbunden die Einführung eines Service Desk als Single Point of Contact sowie die Umsetzung der Prozesse mit entsprechenden ITSM-Systemen. Um mehr Transparenz bezüglich Zuständigkeiten, Nutzungsvoraussetzung und Abhängigkeiten des IT-Dienstleistungsangebots zu schaffen, ist ein aktueller und umfassender Dienstleistungskatalog unabdingbar. Allerdings ist die Darstellung der IT-Angebote stark an technische Dienstbeschreibungen angelehnt und aus Sicht der Hochschulleitung schwer nachvollziehbar. In dem vorliegenden Beitrag wird mit Hilfe einer IT-Landkarte für Dienste die notwendige Transparenz für eine effiziente Kommunikation zwischen Leitungsebene und IT-Organisation vorgestellt.

Keywords: IT-Servicemanagement, IT-Dienste, Prozesslandkarte, Governance

1 Einleitung

MOOCs (Massive Open Online Courses), Lernmanagementsysteme wie moodle oder mahara zur Unterstützung von ePortfolios sowie sinkende Präsenzanteile in der Lehre insbesondere in Weiterbildungsstudiengängen verlangen verlässliche IT-Systeme und Dienstleistungen. Hochleistungsrechnen, Forschungsinfrastrukturen und eine stärkere Vernetzung von Wissenschaftlern sollen durch IT vorangetrieben werden. Die Unterstützung der Verwaltung, z. B. Studierendenverwaltung, Personalverwaltung und Finanzverwaltung, sowie die zunehmende Digitalisierung erfordern eine leistungsfähige und komfortable IT-Unterstützung. Durch derartige gestiegene Anforderungen in allen Bereichen einer Hochschule an IT-Systeme und an Hochschulrechenzentren als deren Betreiber wird die Hochschul-IT zum Getriebenen. Finanzielle Zwänge und regulatorische Vorgaben (Gesetze, Richtlinien und Vorschriften, z. B. Bundesdatenschutzgesetz, Umgang mit SEPA) stellen zudem weitere Herausforderungen dar.

Hochschulrechenzentren stehen vor der Aufgabe, Hochschulprozesse bestmöglich zu unterstützen, was mit der Einführung von IT-Servicemanagement (ITSM) gelingen soll.

[1] Ernst-Abbe-Hochschule Jena, Fachbereich Wirtschaftsingenieurwesen / CIO, Carl-Zeiss-Promenade 2, 07745 Jena, Christian.Erfurth@fh-jena.de
[2] Friedrich-Schiller-Universität Jena, Universitätsrechenzentrum, Am Johannisfriedhof 2, 07743 Jena, Ivonne.Erfurth@uni-jena.de

Das Portfolio der angebotenen IT-Dienste soll einen Mehrwert für die Anwender beispielsweise in Form einer Produktivitätssteigerung oder einer Verringerung der Einschränkungen, denen der Anwender unterliegt, erbringen. Dabei durchlaufen Hochschulrechenzentren einen Wandel weg von der reinen Informationstechnik hin zur Kunden- und Serviceorientierung. Die Aufgaben eines IT-Dienstleisters gehen heutzutage weit über Infrastrukturbetreuung, wie Installation und Pflege von Hardwarekomponenten, sowie deren Verkabelung oder Systemadministration hinaus. Rechenzentren müssen sich zunehmend mit der Individualisierung und Anpassung von Applikationen auf spezielle Prozesse der Hochschule beschäftigen. In einem Campus Management System müssen beispielsweise Prüfungs- und Studienordnungen mit den dazugehörigen Prozessen umgesetzt werden und im Identity Management muss der Lifecycle von Studierenden und Mitarbeitern abgebildet sowie entsprechende Rechtemodelle hinterlegt werden. Des Weiterem werden neue Dienstangebote eingefordert z. B. Forschungsdatenverwaltung oder der Zugang mit mobilen Endgeräten zu bisher internen Applikationen. Diesen Wandel der Hochschulrechenzentren spiegelt die Entwicklung der Empfehlungen der Kommission für IT-Infrastruktur (KfR) der Deutschen Forschungsgemeinschaft (DFG) zur Informationsverarbeitung an Hochschulen über die Jahre wider [DF96], [DF10], [DF14].

Für Hochschulrechenzentren bedeutet dies, das bisherige IT-Angebot im Kontext der Hochschulprozesse zu betrachten, um geeignet Anpassungen und Erweiterungen mit einem möglichst passgenauen IT-Dienstportfolio zu erreichen. Prozessdokumentationen als Basis dafür entstehen an vielen Hochschulen. Insbesondere im Rahmen von ERP-Einführungs- bzw. Umstellungsprojekten werden Prozessdokumentationen vorangetrieben. Mit Hilfe der bereits vorliegenden Hochschulprozesse in den Bereichen Forschung, Studium & Lehre und Verwaltung sowie mit Hilfe von ITSM-Rahmenwerken besteht für Hochschulrechenzentren die Möglichkeit, IT-Dienste zur Unterstützung der Arbeit von Lernenden und Lehrenden sowie für die Verwaltung anzubieten.

Dennoch stellen sich die Fragen, wie einerseits eine Hochschule den Prozess, Anforderungen zu erkennen und zu ermitteln, geeignet aufstellen kann und andererseits wie das Angebot an gebotener IT-Unterstützung den potentiellen Nutzern geeignet übermittelt werden kann [EE14]. Prinzipielle Antworten geben bereits Rahmenwerke wie beispielsweise die IT Infrastructure Library (ITIL) [CHR07] aus dem ITSM oder COBIT [IS14] für die Governance. Die Hochschulen, insbesondere die Rechenzentren, tauschen sich in der Adaption derartiger Rahmenwerke schon seit längerem bilateral oder in Arbeitsgruppen, z. B. im Verein Zentren für Kommunikation und Informationsverarbeitung e. V. (ZKI), aus. Die organisatorischen Strukturen in Hochschulen erschweren teilweise die erfolgreiche Einführung (siehe [KGV12]). In einigen Hochschulen wurden wegen der gestiegenen Bedeutung der IT die Strukturen z. B. durch Etablierung von CIOs oder CIO-Gremien angepasst (siehe ZKI-Studie zu CIOs [Hey14]).

Die Thüringer Hochschulen wurden im vergangenen Jahr vom Ministerium aufgefordert,

I&K Strukturkonzepte vorzulegen. Zur Erhöhung der Vergleichbarkeit der I&K Strukturkonzepte orientierten sich alle Konzepte der Thüringer Hochschulen an einer Vorgabe des ZKI aus dem Jahre 2008 [ZK08]. Weiterhin im Fokus lag eine systematische Aufnahme der Dienstleistungen in einem Katalog. In diesem Kontext wurden eine IT-Landkarte und eine systematische Dienstbeschreibung erstellt. Die Ergebnisse werden in diesem Paper präsentiert.

2 Transparenz des IT-Dienstleistungsangebots

Im Vordergrund bei der Einführung von ITSM an Hochschulrechenzentren stehen die Ziele mehr Transparenz und Kundenorientierung zu schaffen, die Kommunikation zwischen Anwender (Dienstnehmer) und Hochschulrechenzentren (IT-Diensterbringer) sowie innerhalb der IT-Organisation(-en) zu verbessern und, nicht zuletzt, die Qualität bei gestiegenen Anforderungen zu halten bzw. zu verbessern. Durch eine initiale Erfassung und Dokumentation der IT-Leistungen und der Erstellung eines IT-Dienstleistungskatalogs sind erste Grundlagen für mehr Transparenz und verbesserte Kommunikation geschaffen. Dennoch genügt es nicht die Dienste aus technischer Sicht zu betrachten und zu beschreiben, denn auch die Kundenorientierung muss weiter ausgeprägt werden. Insbesondere ist die Schaffung von Transparenz für die Hochschulleitung von Bedeutung, um effektiv Vorgaben für die IT zu definieren wie zum Beispiel Festlegungen zu kritischen Diensten (Business Critical Functions) und Notfallplänen. Die Führung der IT schafft sich durch die Transparenz eine passende Basis für die gezielte Unterstützung und Weiterentwicklung der IT. Wie kann nun eine Verknüpfung der in den jeweiligen Hochschulen vorliegenden Hochschulprozesse mit entsprechenden IT-Diensten erfolgen, um die notwendige Transparenz zu schaffen?

2.1 Zusammenhang zwischen IT-Diensten und Hochschulprozessen

ITIL als „best practice“ für IT-Servicemanagement beschreibt eine Methode zur Entwicklung von neuen IT-Diensten bis hin zum operativen Betrieb des Dienstes, inklusive Wartung, Support und kontinuierliche Qualitätssicherung (Service-Lifecycle). Als Framework fast ITIL zusammen „WAS“ notwendig ist, wenn ein Dienst geplant, entwickelt und betrieben wird (Rollen, Prozesse, etc.). Für die Planung und den Betrieb eines Dienstes stellt ITIL eine enge Verbindung zwischen dem (IT-)Dienst (Business Services), der einen bestimmten Geschäftsprozess (Business Process) des Kunden unterstützt, her [CHR07] (business view). So wird die Hochschullehre als Kernprozess beispielsweise über ein Lernmanagementsystem unterstützt, das eigenständiges Lernen unabhängig von Ort und Zeit fördert. Aus der prozessualen Sicht spielt das konkrete IT-System auf dieser Ebene (noch) keine Rolle. Bei der Planung und Realisierung des IT-Dienstes wird auf Basis der Anforderungen ein passendes IT-System gewählt. Dienste werden wie in diesem Beispiel mit entsprechenden Applikationen und IT-Systemen innerhalb der existierenden Infrastruktur realisiert (technical view, siehe Abb. 1).

Das Kundenmodell (Customer) von ITIL ist nicht direkt innerhalb einer Hochschule umsetzbar (siehe hierzu auch [KGV12]). Aus Sicht von ITIL ist der Kunde der Geldgeber und gestaltet das Angebot an Diensten erheblich mit. „Aufgrund der komplexen Stakeholderstruktur ist es auch für einen internen IT-Dienstleister oft unklar, wer was beauftragen kann bzw. darf. Diese komplexen Strukturen führen häufig dazu, dass sich die Anwender in der Rolle des Kunden wähnen, oder der IT-Dienstleister eigenmächtig Vorgaben für eine IT-Strategie definiert." [KGV12] Daher schlagen wir vor, den Kunden als Geldgeber zu vernachlässigen und eine Differenzierung zwischen (End-) Anwender und Mandant (fachlicher Abnehmer[3]) vorzunehmen (Business Function). In Kooperationsmodellen zwischen Hochschulen, wie es auf Landesebene oftmals üblich ist, ist ein Einsatz vom ITIL-Kundenmodell und weiteren ITIL-Prozessen, wie beispielsweise das Service Level Management, durchaus denkbar und empfehlenswert.

Mit Hilfe der bereits dokumentierten Hochschulprozesse besteht die Möglichkeit, entsprechende Dienste zu identifizieren und zu etablieren sowie entsprechende Technologien (Applikationen) für den Dienst auszuwählen oder zu entwickeln und in die existierende Infrastruktur einzubetten. Auf unterer Ebene (Technische Sicht) sind technische Beschreibungen und Dokumentationen in Hochschulrechenzentren bereits gut ausgeprägt und etabliert.

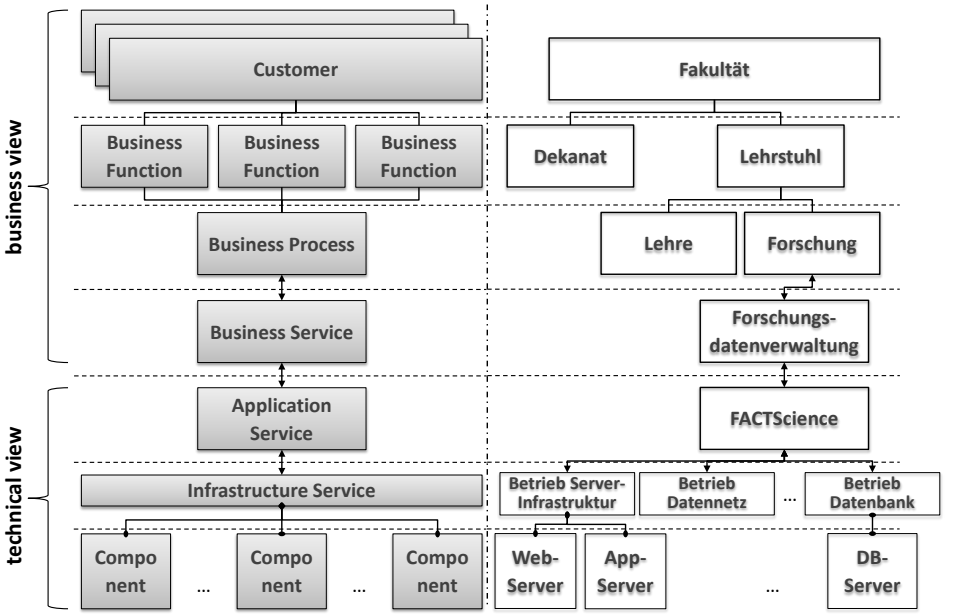


Abb. 1: Sicht auf IT-Dienste inkl. Bsp. – Komponenten und Beziehungen (adaptiert von [CHR07])

[3] Im Falle einer Mittelzuweisung, z. B. durch eine Professur, kann der Geldgeber als fachlicher Abnehmer angesehen werden.

Eine Lücke besteht jedoch insbesondere in der Verknüpfung zwischen Application Service und Business Service hin zur oberen Ebene (Business-Sicht). Mit dem Schließen dieser Lücke zwischen Business- und Technischer Sicht, kann eine Kommunikationsbrücke zwischen Hochschulleitung und IT-Organisation durch mehr Transparenz bezüglich Angebote und Leistungen geschlagen werden. Die Kommunikationsbrücke ist entscheidend für die Governance – also die Führung – nicht nur in der IT sondern für die Hochschule insgesamt. Strategische Entwicklungen (die eine Relevanz für die IT haben) können unter Betrachtung der Risiken fundierter bewertet und entschieden werden. Natürlich muss dies durch geeignete Organisationsstrukturen unterstützt werden, die im Hochschulkontext meist in Form von Gremien abgebildet sind.

2.2 Transparenz der IT-Angebote und Leistungen durch eine IT-Landkarte

Wie kann die Lücke zwischen Business- und Technischer Sicht im Kontext von Hochschulen geschlossen werden und somit mehr Transparenz der IT-Angebote und Leistungen gegenüber einerseits dem (End-)Anwender und andererseits der Hochschulleitung erzeugt werden? In vielen Hochschulen existiert eine Prozesslandkarte, die einen Einstieg in die Modellierung und die Hauptprozesse auf strategischer Ebene der Organisation in einem Wertschöpfungsdiagramm darstellt [Po85]. Oftmals erfolgt eine Dreiteilung in Managementprozesse (Führungsprozesse), Kernprozesse (primäre Prozesse) und unterstützende Prozesse (Supportprozesse, sekundäre Prozesse) [BK00].

Typische Kernprozesse der Hochschule auf oberster, strategischer Ebene sind Studium, Lehre und Weiterbildung sowie Forschung und Transfer. Ein Beispiel für eine verallgemeinerte Prozesslandkarte gibt [Ko07]. In einigen Geschäftsbereichen, insbesondere in der Hochschulverwaltung, sind vereinzelt Prozesse auf tieferen Ebenen dargestellt. Im Arbeitskreis Campus Management des ZKI e.V. hat sich aufgrund der zunehmenden Bedeutung von Hochschulprozessen im letzten Jahr die Arbeitsgruppe Prozess AG etabliert. Die Arbeitsgruppe hat sich als Zielsetzung die Erstellung einer hochschulübergreifenden Prozesslandkarte für Campus Management gesetzt sowie die Entwicklung und Etablierung einer gemeinsamen Teilprozessstruktur. Die Diskussionen im Arbeitskreis verdeutlichen einerseits den wachsenden Bedarf an Prozessdokumentation, zeigen andererseits den Aufholbedarf der Hochschulen in diesem Bereich auf. Detaillierte Prozessmodellierungen sind derzeit im Bereich Campus Management im Entstehen.

Mit der Komponentendarstellung aus Abbildung 1 und den unterschiedlichen, grob dargestellten Prozessbereichen einer Hochschule ist es möglich, die bisher auf technischer Ebene angebotenen IT-Dienste eines Hochschulrechenzentrums überblicksartig in eine IT-Landkarte der Dienste als Business-Sicht zu überführen und somit geschickt in den Kontext der Prozessbereiche zu setzen. Abbildung 2 präsentiert

eine solche, beispielhafte IT-Landkarte, in der die Business-Sicht mit der Technischen Sicht verknüpft wurde. Die Elemente der IT-Landkarte sind Aggregationen, so dass eine Zuordnung überblicksmäßig sichtbar wird. Die oberste Ebene – die Hochschulprozessebene – stellt die Hochschulprozesse lt. Prozesslandkarte mittels noch sehr grob gehaltenen Kernprozessen dar. Mit Hilfe dieser Ebene erfolgt eine erste Verknüpfung mit IT-Diensten – aus Business-Sicht. So wird beispielsweise zur Unterstützung aller Kernprozesse ein zentraler E-Mail-Dienst bereitgestellt. Für die Unterstützung von Lehre, Studium und Weiterbildung beispielsweise Räume mit Rechentechnik oder digitale Medien und Werkzeuge zur Unterstützung von Lehre- und Lernprozessen bereitgestellt.

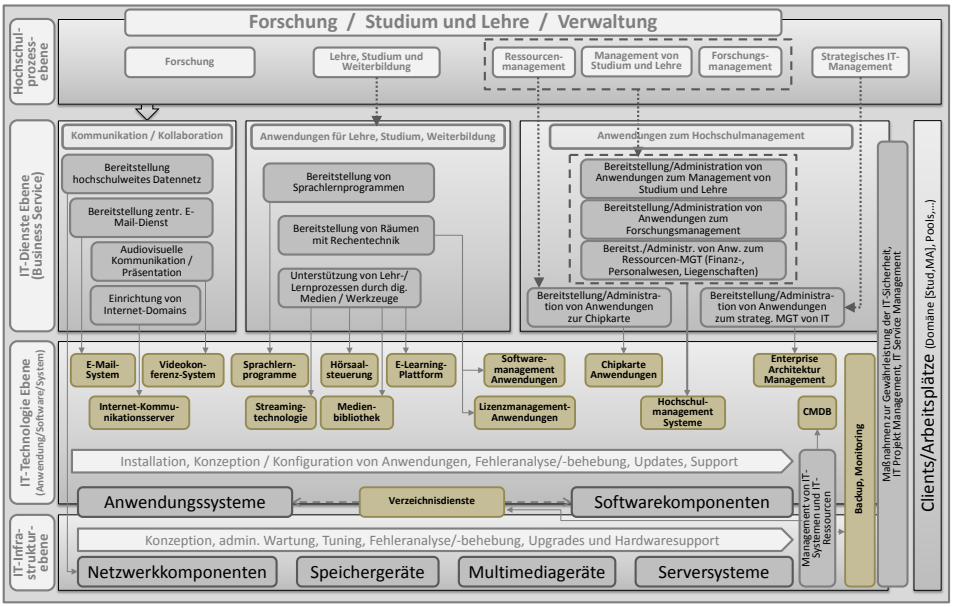


Abb. 2: Auszüge einer IT-Landkarte für Hochschulen

Die Elemente auf der 2. Ebene – IT-Dienste Ebene – sind sogenannte Kerndienste. Kerndienste decken Basisanforderungen ab und werden in der Regel durch weiterführende Anforderungen um erweiternde Dienste ergänzt (siehe Abbildung 3). Erweiterte Dienste sind spezielle Ausprägungen eines Kerndienstes, die beispielsweise für eine eingeschränkte Benutzergruppe oder mit speziellen Qualitätsmerkmalen angeboten werden. Auf dieser Ebene werden die Dienste für die Anwender beschrieben. Derzeit erfolgt die Erstellung vollständiger Servicebeschreibung für den Endanwender und Kurzbeschreibungen des Dienstes zur Web-Darstellung.

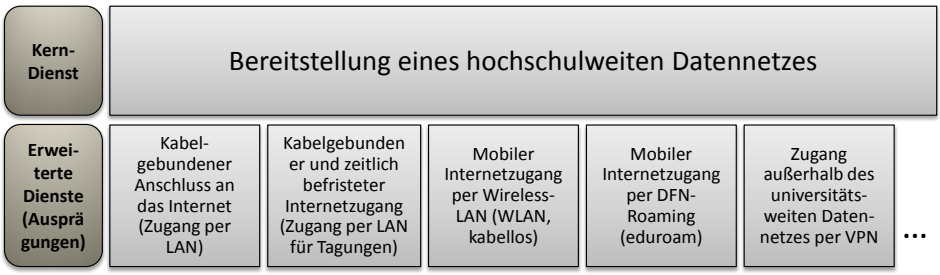


Abb. 3: Kerndienst und deren Ausprägungen (erweiterte Dienste)

Zur besseren Übersichtlichkeit wurden die Kerndienste noch untergliedert in Bereiche, wie [1] Kommunikationsdienste und Kollaboration, [2] Anwendungen für Lehre, Studium und Weiterbildung oder [3] Anwendungen zum Hochschulmanagement. Je nach weiteren Dienstangeboten der Hochschule können noch Kerndienste in den Bereichen [4] Mediadienste, [5] Druckdienste oder [6] Bereitstellung von IT-Ressourcen für Forschung, Lehre und Verwaltung vorliegen. Zusätzlich zu diesen Bereichen kommt ein Bereich für [7] Basis- und Querschnittdienste, in dem Kerndienste wie Maßnahmen zur Gewährleistung der IT-Sicherheit oder Management von IT-Systemen und IT-Ressourcen hineinfallen (siehe Abb. 2 rechts: Dienste, die mehrere Ebenen schneiden). Mittels dieser Bereiche [1-7] wurde an unseren Einrichtungen ein übersichtlicher Dienstleistungskatalog mit 4-8 Kerndiensten pro Bereich erstellt.

Die IT-Dienste werden durch IT-Lösungen auf IT-Technologieebene (3. Ebene von oben) realisiert. Diese Trennung ist durchaus vorteilhaft, da einerseits Dienste auf unterschiedliche Art realisiert werden können, was damit zum Ausdruck gebracht wird, und andererseits eine Technologie in unterschiedlichen Kontexten genutzt werden kann. Somit wird die Rolle und Bedeutung der Technologie klarer und der Blick auf nichtfunktionale Anforderungen wie Verfügbarkeit und Wiederherstellungszeit geschärft. Bei der Definition von Notfallplänen für kritische Dienste (Business Critical Functions) und bei Sicherheitsbetrachtungen ist dies notwendig ebenso wie zur Klärung organisatorischer Fragen im IT-Management. Die in dieser Ebene dargestellten Elemente sind bewusst abstrahiert. In der realen IT-Landkarte sind an dieser Stelle zusätzlich die am Rechenzentrum konkret eingesetzten Technologien angegeben. So kommen beispielsweise als Hochschulmanagementsysteme HIS, DoSV oder SAGE-Lohn oder als E-Learning-Plattform Metacoon zum Einsatz. Für die Technologien existieren entsprechende technische Beschreibungen, wie Betriebskonzept, Notfallplan und Sicherheitskonzept.

Auf unterster Ebene wird lediglich stark abstrahiert angedeutet, dass hinter den IT-Lösungen entsprechende Infrastruktur notwendig ist, um die Technologien zu betreiben. Auf der IT-Infrastrukturebene sind weitere eher technische Dokumentationen notwendig, die beispielsweise in einer Configuration Management Database (CMDB) [AX13]

gepflegt werden können.

2.3 Auf dem Weg zu einer umfassend dokumentierten Dienstbeschreibung

Für eine umfängliche Dienstbeschreibung ist es notwendig den Dienst einerseits so zu beschreiben, dass das Angebot und die Leistungen des Hochschulrechenzentrums, aber auch die Einschränkungen und Zugangsvoraussetzungen eines Dienstes dem (End-) Anwender kommuniziert werden kann (Business-Sicht). Andererseits besteht die Notwendigkeit den Dienst und Zusammenhänge zu anderen Diensten oder Infrastrukturkomponenten technisch zu dokumentieren (Technische Sicht). Eine Dokumentation der Dienste aus technischer Sicht ist in den meisten Hochschulrechenzentren erfolgt. Hierzu gehören beispielsweise Notfallplan, Betriebskonzept und Systemarchitektur, Sicherheitskonzept, ggf. Verfahrensverzeichnis und Einweisungen für den IT-Service Desk sowie, falls interne Dienstleister oder externe Lieferanten in einem Dienst involviert sind, entsprechende Operational Level Agreements (OLA) oder Underpinning Contracts (UC) (siehe Abb. 4).

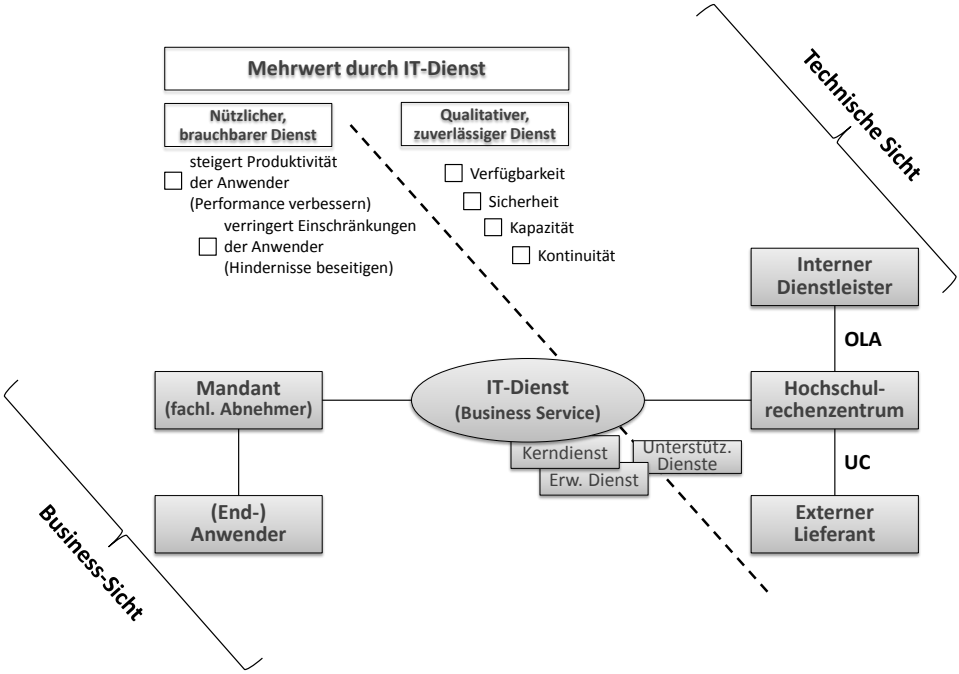


Abb. 4: Überblick für die IT-Dienstbeschreibung – Technische und Business-Sicht (adaptiert von [CHR07])

Die Dienstbeschreibung für den Anwender sollte eine *Kurzbeschreibung* von ein bis max. zwei Sätzen sowie eine *ausführliche Beschreibung* für den (End-)Anwender enthalten. Die ausführliche Beschreibung gibt die Funktionalität des Dienstes ohne Verwendung technischer Fachbegriffe an. Die Beschreibung kann auch über entsprechende Hochschulprozesse erfolgen, falls es zur Verständlichkeit der Funktionalität beiträgt. Zudem sollte der Nutzen des Dienstes beschrieben werden: Was geht schneller, besser, billiger?

Die Angabe von *Schlagwörtern* vereinfacht u.a. eine Umsetzung des Dienstes in entsprechende ITSM-Systeme und ermöglicht dem Anwender eine effiziente Suche nach geeigneten Diensten oder, bei offenen Fragen, in der FAQ. Zusätzlich ist die Angabe *möglicher Nutzer* (für wen ist der Dienst interessant oder nützlich) und / oder *möglicher Einsatzgebiete* (in welchen Bereichen kann der Dienst eingesetzt werden) sowie *Informationen zur Bereitstellung / zum Zugang* in der Dienstbeschreibung erforderlich. Optional kann eine Dokumentation der häufigen Fragen zum IT-Dienst (*FAQ*) angehangen werden einschließlich Hinweisen, wer den Dienst bereits nutzt oder wo der Dienst angesehen bzw. getestet werden kann, bevor er in Anspruch genommen wird (*Referenzen*).

3 Einsatz der IT-Landkarte im Kontext von Servicemanagement und Governance

ITIL beschreibt einen Lebenszyklus für wertbringende IT-Dienste: Von der Strategie (Service Strategy) über Planung (Service Design) bis hin zur Überführung (Service Transition) in den Betrieb (Service Operation). Die IT-Landkarte kann unterstützend für die Aufgaben in den Phasen eingesetzt werden.

Strategie: Ausgangsbasis für die Herleitung von Anforderungen an bestehende und neue IT-Dienste sind die Hochschulprozesse. Je nachdem, auf welcher Detailebene die Prozesse beschrieben sind, können erste Anforderungen an eine IT-Unterstützung abgegriffen werden. Wichtig ist, dass Prozessverantwortliche benannt sind, die bei der Anforderungsermittlung eine wesentliche Rolle spielen: Sie sind die fachlich Beitragenden in der Runde der Stakeholder. Bei unzureichend beschriebenen Prozessen können die Prozessverantwortlichen die fachlichen Details ggf. für die weitere Planung der Umsetzung liefern. In der Strategiephase ist die Wegrichtung der Hochschule durch Ableitung des IT-Konzeptes und von Zielen aus dem Hochschulkonzept wichtig. Wenn das gelingt, zieht die IT in dieselbe Richtung wie die Hochschule selbst und gibt damit optimale Unterstützung. Mit Hilfe der IT-Landkarte kann die Relevanz der IT-Dienste zur Erreichung der Ziele eingeschätzt und festgelegt werden. Selbst Defizite im Serviceangebot lassen sich auf dieser Ebene erkennen. Durch die Abstraktion in Kerndienste und der Zuordnung zu Aufgabenbereichen/Prozessen wird die Einbindung geeigneter Stakeholder unterstützt. Notwendige Änderungen am Serviceangebot lassen

sich in beide Richtungen zwischen Hochschulleitung und IT-Organisation transparenter kommunizieren und bewerten.

Planung: Für neue IT-Dienste aber auch für Änderungen an bestehenden IT-Diensten (Changes) ergeben sich eine Reihe von Fragestellungen (fachliche, technische und organisatorische). Die IT-Landkarte kann dazu beitragen, Lösungsalternativen und Abhängigkeiten im Blick zu behalten und damit in diesem Prozess wieder die geeigneten Stakeholder einzubinden. Betrachtungen zu Risiken werden durch die IT-Landkarte umfassender ermöglicht. In dieser Phase entsteht die Beschreibung der Business-Sicht.

Überführung: Bei der Einbettung des IT-Dienstes in die technische Landschaft bekommen die IT-Verantwortlichen durch die IT-Landkarte und die Beschreibung des Dienstes aus der Business-Sicht einen besseren Einblick in den fachlichen Kontext. Das sorgt für mehr Transparenz und Verständnis für die Relevanz des Dienstes.

Betrieb: Bei der Störungsbearbeitung (Incident Management) und beim Erkennen von Problemstellen (wiederkehrende Störungen – Problem Management) kann die Landkarte zur Ursachensuche und Folgenabschätzung als ergänzendes Hilfsmittel eingesetzt werden. Die Einordnung von Störungsmeldungen aber auch von Service Requests wird für Nutzer durch verständliche Kategorien (z. B. Kerndienst-Bezeichnungen) vereinfacht. ITSM-Werkzeuge können mit Unterstützung der Landkarte eine abgestimmte Konfiguration (Kategorien, Klassifikation, Priorität) erhalten. Für die Nutzer bieten neben der IT-Landkarte auch die Dienstbeschreibungen eine sehr gute Orientierung: Möglichkeiten und Grenzen von Dienstangeboten sind transparent, so dass auch durch Vermeidung falscher Erwartungen eine höhere Zufriedenheit mit den IT-Angeboten erzielt werden kann.

Governance: Die Führung der IT wird durch die Strategie-Phase des ITSM gestützt, in der die IT-Landkarte aber auch verständliche Dienstbeschreibungen eine Rolle spielen. Zudem ist die Bildung geeigneter Organisationsstrukturen wichtig, um Verantwortlichkeiten zu definieren und Steuerungsmöglichkeiten zu ermöglichen. Aus Sicht der IT ist die Einbindung von geeigneten Stakeholdern (Anspruchsgruppen) zur Ausrichtung der IT-Dienste und zur Beurteilung des Wertbeitrages von IT-Diensten essentiell. Das sorgt ebenso für Transparenz in der Hochschule bei allen Beteiligten und kann in Form von geeignet zusammengesetzten Gremien in typische Hochschulstrukturen eingebracht werden.

4 Zusammenfassung

Konzepte des ITSM werden von Hochschulrechenzentren verstärkt angenommen, um sich den gestiegenen Anforderungen aufgrund der umfänglichen Durchdringung der IT in den Hochschulen zu stellen. An den Thüringer Hochschulrechenzentren ist ITSM noch im Aufbau. Die im Rahmen der I&K Strukturkonzepten erstellte IT-Landkarte und der IT-Dienstleistungskatalog hat sich in Gesprächen mit der Hochschulleitung initial

bewährt. Das Aufgabenspektrum der Rechenzentren konnte mit höherer Transparenz der Hochschulleitung vermittelt werden. Insbesondere der Zusammenhang zu den Hochschulprozessen bzw. zu den Aufgabenbereichen der Hochschule und nicht die Aufzählung technischen Systeme, die ein Rechenzentrum betreut, legte dar, in welchem Umfang die IT die Hochschule unterstützt.

In Zukunft muss die praktische Umsetzung erweisen, ob sich die hier präsentierten Ideen und Konzepte etablieren lassen. Der Kommunikationsbedarf von Anwendern und der Hochschulleitung mit der IT-Organisation ist gestiegen, um das Portfolio der IT-Dienste stärker an den Bedürfnissen und der Wegrichtung einer Hochschule auszurichten. Zur Unterstützung der Kommunikation und zur besseren Planung ist die Etablierung dieser IT-Landkarte für Dienste empfehlenswert, die im Kontext von ITSM wertvolle Verwendung finden kann. Insbesondere in der Strategiephase kann die Landkarte für die Weiterentwicklung und Standardisierung der IT beitragen. Des Weiteren kann diese für die Governance von Bedeutung sein: Der Einsatz als Kommunikationsbasis in etablierten Gremien hilft u.a. Entscheidungen fundiert herbeizuführen. Bei der Einführung von ITSM legen die Betrachtungen zum Aufbau der Landkarte und die weiterführenden Beschreibungen der IT-Dienste auf Business-Sicht eine gute Ausgangsbasis. In Prozessen wie beispielsweise Change-Management kann diese Sicht gewinnbringend genutzt werden. Erste Erfahrungen im Rahmen der erstellten I&K Strukturkonzepte zeigen, dass diese Sichtweise viel Potential hat.

Literaturverzeichnis

[AX13] AXELOS: ITIL Glossar und Abkürzungen Deutsch. 2013. Online: https://www.axelos.com/Corporate/media/Files/Glossaries/ITIL_2011_Glossary_DE-v1-2.pdf, 12.12.2014.

[BK00] Becker J., Kahn D.: Der Prozess im Fokus: In Becker J., Kugeler M., Rosemann M. (Hrsg.): Prozessmanagement – Ein Leitfaden zur prozessorientierten Organisations-gestaltung, Springer, Berlin Heidelberg, 2000.

[Bu99] Burrow, J.: Going beyond labels: A framework for profiling institutional stakeholders. Contemporary Education, vol. 70(4), pp. 5–10, 1999.

[CHR07] Cartlidge, A.;Hanna, A.; Rudd, C.; Macfarlane, I. et.al: An introductory overview of ITIL V3. The UK Chapter oft he itSMF, London, 2007.

[DF96] DFG: Informationsverarbeitung und Rechner für Hochschulen 1996-2000. Empfehlung der Kommission für Rechenanlagen, 1996.

[DF10] DFG: Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme. Empfehlung der Kommission für IT-Infrastruktur für 2011-2015, 2010.

[DF14] DFG: Veröffentlichungen der Kommission für IT-Infrastruktur. Online: http://dfg.de/dfg_profil/gesamtliste_publikationen/it_infrastruktur/index.html, 12.12.2014.

[EE14] Erfurth, C.; Erfurth, I.: Towards business alignment of IT services in universities: Challenges in elicitations of requirements for IT services. IEEE 1st International Workshop on the Interrelations between Requirements Engineering and Business Process Management (REBPM), pp.11,14, 25-25 Aug. 2014.

[FS08] Fine, B.; Schneider, N.: Service catalog management – Best practices and best practical advice. The UK Chapter oft he itSMF, London, 2007.

[Hey14] von der Heyde, M.: CIOs und IT-Governance an deutschen Hochschulen. Lang, U.; Wimmer, M. (Hrsg.), ZKI e.V., 2014. Online: www.zki.de/publikationen/, 08.06.2015.

[IS14] ISACA: COBIT 5: A business framework for the governance and management of enterprise IT. Online: http://www.isaca.org/COBIT, 28.10.2014.

[KGV12] Knittl, S.; Grindler, A.; Vellguth, K.: IT-Servicemanagement Rahmenwerke - wie sie sinnvoll in Universitäten einsetzbar sind. 5. DFN-Forum Kommunikationstechnologien. GI-Edition - Lecture Notes in Informatics (LNI), P-203, Bonner Köllen Verlag, 2012; S. 85-94.

[Ko07] Kocian, C.: Prozesslandkarte für Hochschulen. Die neue Hochschule, vol. 2. , hlb, Bonn, 2007.

[Po85] Porter, M.E.: Competitive Advantage: Creating and Sustaining Superior Performance. New York 1985.

[ZK08] Zentren für Kommunikation und Informationsverarbeitung e. V.: IuK-Strukturkonzepte für Hochschulen. Empfehlungen zu Zielsetzung, Aufbau und Inhalt. ZKI e.V., 2008. Online: http://www.zki.de/fileadmin/zki/Publikationen/IuK-SK_Gesamt-V1_0_02.pdf, 08.06.2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Integrierte Ressourcen-Planung an Hochschulen

Gerald Lach[1] Mirjana Lach[2] Erhard Zorn[3]

Abstract: Die Ressourcen-Planung bietet ein großes Potential, um die knappen Mittel an Hochschulen effizienter einzusetzen und so die Lehre zu verbessern sowie Einschränkungen zu vermindern. Wir stellen ein integriertes Ressourcen-Management-System vor, das Methoden der mathematischen Optimierung verwendet, um die Klausurterminplanung, Lehrveranstaltungsplanung, Tutorieneinteilung und Personaleinsatzplanung zu bewältigen. Das System haben wir an der RWTH Aachen, der TU Berlin und der TU München eingesetzt; die dabei erzielten Ergebnisse fassen wir zusammen.

Keywords: Ressourcen-Management, Klausurterminplanung, Lehrveranstaltungsplanung, Tutorieneinteilung, Mathematische Optimierung

1 Einleitung

In den letzten Jahren hat der Einsatz integrierter Campus-Management-Systeme stark an Bedeutung gewonnen. Hierzu gibt es ein reges Forschungsinteresse (z. B. [AA10], [BK09]). Viele Hochschulen treiben mit großem Aufwand Projekte voran, um den kompletten Lebenszyklus ihrer Studierenden – von der Bewerbung bis zur Exmatrikulation – zu jeder Zeit elektronisch abbilden und effizient verwalten zu können. Anders als ein integriertes Campus-Management steht der Aufbau eines integrierten Ressourcen-Managements selten im Fokus der Hochschulen. Häufig wird der Nutzen durch ein verbessertes Ressourcen-Management nicht erkannt, oder aus politischen Gründen werden Veränderungen bei der Raum- und Personalverteilung gescheut. Ein effizientes Ressourcen-Management scheint an deutschen Hochschulen überwiegend negativ interpretiert und häufig mit unnötig aufgezwungenen Sparmaßnahmen gleichgesetzt zu werden. Abwehrhaltungen werden indirekt gerne mit der „Freiheit von Forschung und Lehre“ begründet. Ein effizientes Ressourcen-Management kann jedoch dazu beitragen, mit beschränkten Ressourcen einen möglichst optimalen Lehrbetrieb sicherzustellen. Es ist schwer nachvollziehbar, warum bei immer knapperen Haushaltsmitteln eher darüber nachgedacht wird, z. B. Fachgebiete zu streichen, als Kosten durch einen effizienteren Einsatz der Raum- und Personalressourcen einzusparen. Die Freiheit von Forschung und Lehre wird dadurch eher unterstützt als durch Einzelne, die auf ihren „angestammten“ Ressourcen zu Lasten der anderen beharren.

Bei Diskussionen zu Einsparmöglichkeiten stehen in der Regel Personalressourcen an erster Stelle; sie sind der größte Posten im Budget. Raumressourcen, mit den dazu gehörigen Kosten für Energie, Personal wie Hausmeister etc., stellen jedoch den zweitgrößten Posten im Haushalt der Hochschulen dar, werden jedoch oft nicht ausreichend betrachtet.

[1] Technische Universität Berlin, innoCampus, Straße des 17. Juni 136, 10623 Berlin, lach@math.tu-berlin.de
[2] Technische Universität Berlin, innoCampus, Straße des 17. Juni 136, 10623 Berlin, mlach@math.tu-berlin.de
[3] Technische Universität Berlin, innoCampus, Straße des 17. Juni 136, 10623 Berlin, erhard@math.tu-berlin.de

Die Verwaltung knapper Ressourcen ist eine an Hochschulen jedes Semester wiederkehrende Aufgabe. Die manuelle Koordination von Lehrveranstaltungen, Räumen, wissenschaftlichem und studentischem Personal in der Lehre ist eine komplexe Ressourcen-Management-Aufgabe, die von den Hochschulen nur mit großem personellen Aufwand – und häufig nicht zufriedenstellend (siehe z.B. [Op], [FV], [ot]) – bewältigt wird. Die meisten Hochschulen können dabei keine Aussage über die Qualität ihrer Ressourcen-Planung machen. Wichtige Kennzahlen wie die Belegungs- bzw. Auslastungsquoten von Hörsälen, die Anzahl an Teilnehmern einer Veranstaltung oder nur die Anzahl der Veranstaltungen in einem Semester stehen den Hochschulen meist nur in unzureichender Qualität zur Verfügung. Ein Grund ist sicherlich die hohe Komplexität universitärer Ressourcen-Planungen. So stellt nicht nur die Vielzahl der zu beachtenden Nebenbedingungen für die Planer eine Herausforderung dar, auch die simultane Koordination aller Beteiligten – wie z. B. Fachgebiete, Raumverwalter und Studienbüros – ist eine enorme Herausforderung.

In dieser Arbeit stellen wir unseren Ansatz zum Management universitärer Ressourcen dar. Die meisten anderen, uns bekannten, Methoden konzentrieren sich auf eine „Beibehaltung" des Status Quo. Dagegen sind unsere universitären Ressourcen-Management-Methoden darauf ausgelegt, flexibel und frühzeitig auf Änderungen der Rahmenbedingungen reagieren zu können.

In Abschnitt 2 geben wir einen Überblick über die von uns betrachteten Ressourcen-Planungsprobleme, und in Abschnitt 3 beschreiben wir die üblichen Ansätze dafür. In Abschnitt 4 stellen wir unseren Ansatz und in Abschnitt 5 seine Umsetzung samt Ergebnissen in Abschnitt 6 vor.

2 Ressourcen-Management an Hochschulen

Das Ressourcen-Management an Hochschulen umfasst die Verwaltung vieler Ressourcen. Beispiele hierfür sind Räume, Personal, aber auch der Fuhrpark oder Beamer. Wir konzentrieren uns in dieser Arbeit auf innovative Methoden zum Management der Ressourcen *Raum, Lehrpersonal und Zeit*. Wir betrachten die folgenden vier Planungsprobleme:

Lehrveranstaltungsplanung Bei der Lehrveranstaltungsplanung werden die Termine aller periodisch stattfindenden Lehrveranstaltungen (in der Regel ohne Tutorien, dazu siehe unten) konfliktfrei koordiniert. Neben den Studienverlaufsplänen und den Raumressourcen sind dabei die Verfügbarkeiten der Dozierenden sowie zeitliche Abfolgebedingungen zwischen (Teilen der) Lehrveranstaltungen zu berücksichtigen.

Klausurterminplanung Bei der Klausurterminplanung wird jeder Klausur (oder schriftlichen Leistungskontrolle) ein Termin und ausreichend Raumkapazität zugewiesen. Wichtig ist die Berücksichtigung der Studienverlaufspläne aller Studien- und Prüfungsordnungen sowie die Einhaltung bestimmter Abstände zwischen den Klausuren, um den Studierenden eine gute Vorbereitung zu ermöglichen.

Tutorieneinteilung Viele Veranstaltungen bieten neben den Vorlesungen und großen Übungen stoffvertiefende Übungen in Kleingruppen (Tutorien) an. Die Studierenden der Veranstaltungen werden bei der Tutorieneinteilung auf die zugehörigen Tutorien verteilt. Dies geschieht in der Regel am Beginn der Vorlesungszeit.

Personaleinsatzplanung Das wissenschaftliche Personal und die studentischen Beschäftigten, die in Tutorien unterrichten (Tutoren/innen) werden Lehrveranstaltungen zugewiesen. Zu beachten ist dabei, in welchen Lehrveranstaltungen das Personal auf Grund seiner Qualifikation einsetzbar ist.

Die Klausurterminplanung nimmt eine Sonderstellung ein: Hier werden Ressourcen für die *vorlesungsfreie* Zeit geplant, bei den anderen drei Problemen wird für die *Vorlesungszeit* geplant. Die Klausurterminplanung hat daher wenige Abhängigkeiten zu den anderen Planungsproblemen. Die Ergebnisse der Lehrveranstaltungsplanung, Tutorieneinteilung und Personal-Einsatzplanung beeinflussen sich gegenseitig. So führt die Reduzierung der Anzahl der Tutoren/innen bei einer Veranstaltung dazu, dass für diese Veranstaltung weniger, aber größere Räume benötigt werden, da mehr Studierende einem Tutorium zugeordnet werden.

2.1 Ressourcen-Management an Hochschulen in der Forschung

Die aufgeführten Stundenplanungsprobleme werden schon seit langem algorithmisch untersucht. Es gab viele Versuche, seine kombinatorische Struktur zu erforschen und effiziente Lösungsstrategien zu entwickeln. 1996 zeigten Cooper und Kingston, dass es sich beim Stundenplanproblem um ein sogenanntes *NP-schweres* Problem handelt [CK96] (für Details von NP-schweren Problemen siehe [GJ78]). Man kann daher nicht erwarten, einen *„schnellen“* Algorithmus zu finden, der das Problem exakt löst. In den letzten Jahren wurden in der Forschung vor allem heuristische Methoden für das Stundenplanungsproblem untersucht. Dies lässt sich besonders gut verfolgen an den Tagungsberichten zur Konferenz-Serie *Practice and Theory of automated Timetabling (PATAT)*[4] [Bu04, BR06].

Seit 2002 gibt es einen internationalen Wettbewerb zur Stundenplanerstellung (ITC)[5][6] [Mc10], in dem Forschergruppen gegeneinander antreten, um die beste Lösungssoftware zu ermitteln. Dazu wurden standardisierte Probleme veröffentlicht, die in der Community als Benchmarks anerkannt sind. In der Kategorie *University Course Timetabling*, ähnlich der *Vorlesungsplanung*, wurden Probleme mit nicht mehr als 300 Kursen betrachtet. Die dort vorgestellten Algorithmen sind daher oft nicht für den Einsatz an deutschen Hochschulen geeignet. Nur wenige in der Forschung betrachtete Lösungsalgorithmen sind an Probleminstanzen mit einer praxisrelevanten Größe getestet [La07, LL08, MR14].

[4] `http://www.asap.cs.nott.ac.uk/patat/patat-index.shtml`
[5] `http://www.cs.qub.ac.uk/itc2007`
[6] `http://www.idsia.ch/Files/ttcomp2002`

Untersuchungen zu effizienten Arbeitsablaufplanungen beim Ressourcen-Management an Hochschulen sind uns nur wenige bekannt. In [Go10] wurden Herangehensweisen zur effizienten Umsetzung der Tutorieneinteilung vorgestellt.

2.2 Herausforderungen für das Ressourcen-Management

An den Hochschulen sind die Rahmenbedingungen für das Ressourcen-Management ständigen Änderungen unterworfen: Prüfungsordnungen ändern sich oder neue kommen hinzu, und Räume werden wegen Sanierungsmaßnahmen gesperrt. Eine besondere Herausforderung war der doppelte Abiturjahrgang mit stark gestiegenen Anfängerzahlen. Insgesamt hat sich der Bolognaprozess sehr stark bei der Stundenplanung bemerkbar gemacht: Die alten Pläne können nicht mehr nur durch wenige Änderungen angepasst werden, weil in zu kurzer Zeit zuviele Änderungen nötig geworden sind. Dabei können bereits wenige Änderungen zu großen Schwankungen bei den Teilnehmerzahlen führen, und das wirkt sich u. a. auf die Tutoreneinsatzplanung aus. Dieses Beispiel zeigt die hohe Komplexität des Ressourcen-Managements an Hochschulen.

Neben der planerischen Komplexität stellt die Einbeziehung aller Beteiligten eine Herausforderung dar. Alle Stakeholder müssen von der Bedeutung einer guten Ressourcen-Planung überzeugt werden. Gleichzeitig muss ihnen vermittelt werden, dass im Sinne einer transparenten und gerechten Ressourcen-Verteilung nur gut begründete Ausnahmen möglich sind, wenn es etwa um die Präferenzen bei den zu nutzenden Ressourcen geht.

Flexibilität ist daher eine der wichtigsten Anforderungen an das Ressourcen-Management, um auf wechselnde Rahmenbedingungen eingehen zu können. Manche Änderungen sind lange im Voraus bekannt (z. B. der doppelte Abiturjahrgang), aber es gibt auch kurzfristig nötige Anpassungen (z. B. auf Grund einer Raumsperrung).

Kurzfristige Änderungen der Rahmenbedingungen Bei kurzfristig notwendigen Anpassungen im aktuellen Ressourcen-Management werden häufig nur wenig zufriedenstellende Lösungen gefunden. Beispiele hierfür sind:

- Nicht vorhersehbare Schwankungen der Studierendenzahlen je Veranstaltung
- Kurzfristige Kündigung studentischer Mitarbeiter/innen
- Raumsperrungen wegen Baumängeln, Raumentzug wegen Sonderveranstaltungen

Mittelfristige Änderungen der Rahmenbedingungen Änderungen, die mindestens 4–5 Monate im Voraus bekannt sind, bezeichnen wir als mittelfristig. Nötige Anpassungen sind in diesem Fall bereits in der Planungsphase bekannt und können in einem flexiblen Ressourcen-Management leicht berücksichtigt werden. Dabei ist es wichtig, die neuen Rahmenbedingungen korrekt zu erfassen. Beispiele hierfür sind:

- Einführung neuer Studien- und Prüfungsordnungen
- Änderung von Studienverlaufsplänen
- Fluktuation beim wissenschaftlichen Personal
- Raumsperrungen wegen Events und geplanter Baumaßnahmen

Langfristige Änderungen der Rahmenbedingungen Im Gegensatz zu kurz- und mittelfristigen Änderungen stehen langfristig zu berücksichtigende Änderungen der Rahmenbedingungen meist im Zusammenhang mit strategischen Entscheidungen der Hochschule. Beispiele hierfür sind:

- Langfristige Gebäudeplanung
- Profil der Hochschule und damit Ausrichtung der Studiengänge
- Langfristige Planung der Zulassungszahlen

3 Verbreitete Methoden des Ressourcen-Managements

Um Ressourcen-Planungsprobleme wie das Stundenplanproblem zu lösen, werden vorwiegend zwei Methoden eingesetzt: Planung „von Hand“ und das *first-come, first-served*-Prinzip. Am vertrautesten ist die Planung „von Hand“; bei einer Software-Lösung werden Planungshilfen geboten, die Angebot und Nachfrage visualisieren oder Konflikte anzeigen. Die Zuordnung von Ressourcen kann durch übliche *Drag & Drop*-Techniken erleichtert werden. Die Software hilft anschließend bei der Information der Beteiligten, z. B. durch Anzeige des individuellen Stundenplanes. Diese Methode bedarf eines hohen Personal- und Zeitaufwandes. Daher besteht das Interesse, einmal bestimmte Lösungen nur noch selten/wenig zu ändern. Für die Lehrveranstaltungsplanung ist dies ein übliches Verfahren; durch die im Laufe der Zeit immer wieder nötigen „kleinen“ Änderungen „wächst“ ein Stundenplan. Die Übernahme der alten Stundenpläne ist daher das Standardverfahren auch bei der HIS-Software LSF, die an vielen Hochschulen in Deutschland eingesetzt wird. Beim „first-come, first-served“-Prinzip werden die Ressourcen zur freien Wahl zur Verfügung gestellt, bis sie erschöpft sind. In der einfachsten Form wird dies an Hochschulen durch Eintragen in Teilnehmerlisten realisiert, fortgeschrittene Web-Lösungen bereiten die Daten digital auf. Dieses Verfahren eignet sich nur, wenn ausreichende Ressourcen zur Verfügung stehen. Die Studierenden als Ressourcen-Nehmer erhalten auch dann in der Regel suboptimale Lösungen, da sie kaum einen Gesamtüberblick über das Angebot haben und ihre Ressourcen nur sukzessive buchen können. Und das Lehrpersonal als Ressourcen-Anbieter kann auf Schwankungen bei der Nachfrage schlecht reagieren. Da die Verteilung der Ressourcen im Allgemeinen zeitlich nicht aufeinander abgestimmt wird, müssen die Ressourcen-Anbieter immer wieder Anpassungen vornehmen. Je später jemand Ressourcen wählt (etwa wegen später Immatrikulation), umso schlechter wird seine Lösung.

Die Ressourcen-Planung mittels des „first-come, first-served“ Prinzips oder mittels der Übernahme aus dem vorhergehenden Semester bietet Planungssicherheit; außerdem ist

dies technisch leicht umsetzbar. Andererseits führt gerade das Prinzip der Datenübernahme zusammen mit der Buchung der Restressourcen nach dem „first-come, first-served“ Prinzip zum Horten von Ressourcen: Aus Angst, nicht genug Ressourcen zu bekommen, buchen die Anwender meist zuviele Ressourcen, die aber anschließend nicht wieder freigegeben werden, um sie in den nächsten Jahren nicht wieder zu verlieren. Als Folge sind trotz ausreichend vorhandener Ressourcen meist keine mehr buchbar.

Die wenigen Software-Lösungen zur automatisierten Ressourcen-Planung an Hochschulen behandeln ”nur“ das Problem der Lehrveranstaltungsplanung. Es sind uns keine Ansätze bekannt für Lehrveranstaltungsplanung, Klausurterminplanung, Tutorieneinteilung und Tutoren-Einsatzplanung in abgestimmten Arbeitsprozessen.

4 Grundidee unseres Ansatzes

Unsere Herangehensweise umfasst drei Schritte:

- **Modellierung:** Ein Ressourcen-Planungsproblem wird mathematisch als *Optimierungsproblem* modelliert.
- **Algorithmusentwicklung:** Es wird ein Algorithmus entwickelt, der das zum Ressourcen-Planungsproblem gehörende Optimierungsproblem löst.
- **Arbeitsprozessentwicklung:** Es werden Arbeitsprozesse entwickelt, die den Einsatz des Algorithmus zum Lösen des Optimierungsproblems vorbereiten und die im Einklang mit den Abläufen an der Hochschule stehen.

Die wichtigste Grundlage unseres Ansatzes ist die Modellierung als Optimierungsproblem: Damit ist ein *objektiver Vergleich* verschiedener Lösungen möglich. In die mathematische Modellierung geht eine sogenannte Kostenfunktion ein, in ihr wird das Maß für die Güte der Lösungen festgelegt. Der Wert der Kostenfunktion einer Lösung kann dabei aus realen Kosten und aus Opportunitätskosten bestehen. Der Algorithmus ermittelt aus den vorliegenden Daten eine Lösung – sofern es eine Lösung gibt –, für die die Kostenfunktion einen minimalen Wert annimmt. Eine solche Lösung heißt *optimal*. Der Algorithmus garantiert, dass keine bessere Lösung, d. h. Lösung mit geringeren Kosten, existiert. Sollte keine Lösung existieren, die allen Anforderungen genügt, wird eine Lösung angegeben, die möglichst viele Anforderungen berücksichtigt. Zusätzlich wird angegeben, welche Anforderungen nicht erfüllt werden konnten. Damit kann ermittelt werden, welche Kombination von Anforderungen die Lösbarkeit verhindern. Außerdem haben wir neue Arbeits- und Kommunikationsabläufe entworfen, die besonders bei der Bereitstellung der benötigten Daten unterstützen, die in dem Umfang nicht digital vorlagen.

5 Umsetzung

Die Ausgestaltung der Arbeits- und Kommunikationsabläufe ist ein zentraler Punkt unserer Arbeit. Gut strukturierte Arbeitsabläufe sowie eine schnelle und freundliche Kom-

munikationspolitik sind eine Grundvoraussetzung, um alle Beteiligten von der Mitwirkung an den Prozessen überzeugen zu können. Ohne die Mitarbeit eines großen Teils der Beschäftigten an der Hochschule ist die Umsetzung nicht möglich. In den folgenden Abschnitten geben wir die wichtigsten Eckpunkte an, die wir bei der organisatorischen, prozessualen und IT-technischen Umsetzung beachtet haben.

5.1 Organisation

Um ein starres Ressourcen-Management zu einem flexiblen umzuwandeln, müssen alte Strukturen aufgebrochen und verändert werden. Entschließt man sich etwa dazu, Tutoren/innen möglichst flexibel in der Lehre einzusetzen – d. h. in allen Veranstaltungen, für die sie qualifiziert sind –, muss zunächst geklärt werden, welcher Organsiationseinheit die Tutoren/innen zugeordnet sind. Hier bietet sich eine übergeordnete Organisationseinheit an; ein Fachgebiet scheint bei einer solchen Anforderung nicht mehr geeignet zu sein.

Bevor ein flexibles Ressourcen-Management eingeführt wird, müssen klare, transparente Regeln zur Ressourcen-Verteilung festgelegt werden. Es müssen etwa folgende Fragen offen diskutiert und die Entscheidungen transparent kommuniziert werden: Gibt es bevorzugte Buchungsrechte bei manchen Räumen? Auf wie viele Tutoren/innen hat eine Veranstaltung Anspruch? Gibt es Veranstaltungen oder Dozierende, für die die Termine fixiert werden? Daraus ergibt sich ein klarer Leitfaden für das Team des Ressourcen-Managements. An der Hochschule muss transparent entschieden werden, ob ein flexibles optimiertes Ressourcen-Management genutzt werden soll, um die Ressourcen und damit Kosten zu sparen, oder ob und wieviel Überkapazität bei den Ressourcen erhalten bleiben soll. Dies kann nur mit Blick auf die Gesamtsituation der Hochschule entschieden werden.

5.2 Prozess

Sind an der Hochschule alle organisatorischen Rahmenbedingungen für ein flexibles Ressourcen-Management festgelegt worden, müssen sie umgesetzt werden. Dazu müssen Arbeitsprozesse implementiert werden, die sicherstellen, dass die Daten in der benötigten Qualität bereitgestellt werden. Eine gute Datenqualität ist eine wichtige Voraussetzung für ein funktionierendes Ressourcen-Management. Alle Beteiligten sollten daher die Informationen bzw. Daten in ihrem Verantwortungsbereich selbst bearbeiten können. Wir haben uns deswegen für dezentral organisierte Prozesse entschieden, die jeweils von einem oder mehreren Prozess-Administratoren begleitet werden. Die meisten Informationen werden von den jeweiligen Wissensträgern selbst eingetragen, beispielsweise kennt das Team des Ressourcen-Managements nicht die persönliche Verfügbarkeiten eines Dozierenden oder dessen Raumpräferenzen. Diese Daten sollten daher von den Dozierenden selbst im System hinterlegt werden. Die Erfahrung hat allerdings gezeigt, dass grundlegende Informationen wie z. B. Studienverlaufspläne von einer zentralen Stelle eingepflegt werden sollten, um eine ausreichende Qualität der Daten gewährleisten zu können.

Der grobe Ablauf der von uns entwickelten Arbeitsprozesse gliedert sich in die folgenden sechs Phasen:

P1: Vorbereitungen In dieser Phase bereiten die Prozess-Administratoren die Planung vor. Lang- und mittelfristige Änderungen, die nicht von den Dozierenden abhängig sind, werden in das Ressourcen-Management-system eingepflegt. Typische Aufgaben sind: Festlegung und Kommunikation der Termine, Festlegung von Sperrungen aufgrund externer Veranstaltungen (Konferenzen, Messen etc.), Einpflegen von Änderungen in den Studienverlaufsplänen.

P2: Dezentrale Dateneingabe Sind die Vorbereitungen abgeschlossen, werden die jeweiligen beteiligten Nutzer aufgefordert, relevante Informationen in das System einzutragen. In der Lehrveranstaltungsplanung sind das beispielsweise die Verfügbarkeiten der Dozierenden und in der Klausurterminplanung z. B. die Wunschtermine der Durchführenden. Die Prozess-Administratoren unterstützen in dieser Phase die Nutzer und verfolgen den Fortschritt der Dateneingabe. Sollten nicht alle Beteiligten wie erwartet ihre Daten im System hinterlegen, müssen die Prozess-Administratoren dafür sorgen, dass die fehlenden Daten im System vervollständigt werden. Am Ende dieser Phase sind alle lang- und mittelfristigen Veränderungen im Ressourcen-Management-System hinterlegt worden, so dass sie in der aktuellen Planung berücksichtigt werden können.

P3: Zentrale Datenbereinigung Nach Abschluss der dezentralen Dateneingabe kontrollieren die Prozess-Administratoren abschließend die Dateneingabe der Nutzer. Das Ressourcen-Management-System muss dazu den Prozess-Administratoren geeignete Analyse-Werkzeuge etwa für Statistiken zur Verfügung stellen.

P4: Plan-Erstellung Sind alle benötigten Daten im System hinterlegt und kontrolliert, wird die Ressourcen-Planung automatisiert durch die Optimierungssoftware durchgeführt und es wird ein Plan erstellt. Die Prozess-Administratoren können hierbei verschiedene Szenarien erstellen mit unterschiedlichen Parametereinstellungen, um dann mehrere unterschiedliche Ergebnisse zu vergleichen. Das am besten geeignete Ergebnis wird am Ende der Plan-Erstellungs-Phase von den Prozess-Administratoren ausgewählt.

P5: Interne Veröffentlichung und Plan-Nachbearbeitung Bevor die erstellte Ressourcen-Planung veröffentlicht wird (z. B. im Online-Vorlesungsverzeichnis), können sie alle aktiv daran Beteiligten einsehen, um gegebenenfalls noch Änderungen vorzunehmen.

P6: Veröffentlichung und Einarbeitung kurzfristiger Veränderungen Nach Ablauf einer Frist wird die zu diesem Zeitpunkt aktuelle Planung übernommen und veröffentlicht. Das Ressourcen-Management-System muss den Prozess-Administratoren erlauben, kurzfristig auftretende Änderungen der Rahmenbedingungen effizient in die vorhandene Ressourcen-Planung einzuarbeiten.

5.3 IT-Umsetzung

Die komplexen Arbeitsabläufe aus Abschnitt 5.2 sind an einer Hochschule nur umsetzbar, wenn sie durch ein IT-System unterstützt werden. Unser IT-System besteht aus zwei Komponenten: *Web-Applikation* zur Datenpflege und *Optimierungstools*.

Web-Applikation Neben den üblichen Anforderungen an eine Web-Applikation – z.B. Browserkompatibilität und ein ansprechendes Design – sollte bei der Entwicklung der Web-Applikation aufgrund der dezentral organisierten Prozesse ein Schwerpunkt auf das Rechtesystem gelegt werden. Es muss zum einen granular genug sein, um die jeweiligen Zugriffsrechte zu steuern; zum anderen muss es so konzipiert sein, dass es von den Prozess-Administratoren effizient verwaltet werden kann.

Optimierungstools Die Optimierungstools sind das eigentlich Kernstück unserer Idee eines optimierten flexiblen Ressourcen-Managements. Hierbei wird eine Reihe von Anforderungen an die Algorithmen gestellt. Die Optimierungsprobleme sind an großen Hochschulen sehr groß und komplex, so dass viele Algorithmen nicht in der Lage sind, Lösungen zu berechnen. Darüber hinaus müssen die Optimierungstools die geforderten, teilweise sehr hochschulspezifischen Rahmenbedingungen, in ein geeignetes Modell zu übersetzen. Besonders in den letzten Jahren gab es in der Forschung viele Bemühungen, geeignete Algorithmen und Modelle zum Lösen der Klausurterminplanung, der Lehrveranstaltungsplanung und der Tutorieneinteilung zu entwickeln. Die meisten dort betrachteten Probleme reichen allerdings nicht an die Größen heran, die in den Planungsproblemen an vielen deutschen Hochschulen auftreten.

6 Ergebnisse

Im Sommersemester 2003 erprobten wir erstmals unseren Ansatz zum Ressourcen-Management bei der Tutorieneinteilung an der TU Berlin. Seitdem haben wir unser IT-System so erweitert, dass nun die Klausurterminplanung, die Lehrveranstaltungsplanung und die Tutorieneinteilung unterstützt werden. Wir haben das System erfolgreich an der RWTH Aachen, der TU Berlin und der TU München eingesetzt. In den folgenden Abschnitten stellen wir kurz unsere Erfahrungen und Ergebnisse dar.

6.1 Klausurterminplanung

Seit Sommersemester 2010 wird unsere Methode zur Koordination der Klausurtermine an der TU Berlin eingesetzt. Wie bei der Tutorieneinteilung war für die Klausurveranstalter die Teilnahme an der Klausurplanung freiwillig. Trotzdem ist die Zahl der teilnehmenden Veranstaltungen jedes Semester gestiegen, weil immer mehr Veranstalter den Service freiwillig in Anspruch nehmen. Einen wichtigen Grund für die große Akzeptanz sehen

wir im geringen Planungsaufwand von 10 bis 20 Minuten je Klausur für die Verantwortlichen, wobei die Überschneidungsfreiheit von Klausuren für alle Studiengänge sichergestellt wird – nervenaufreibende und zeitaufwändige Abstimmungen entfallen. Auch die Steigerung der durchschnittlichen Raumauslastung von 66 % auf 95 % während des Klausurzeitraums in den größten und beliebtesten Räumen und die damit einhergehende verbesserte Raumzuteilung trug sicherlich zur Zufriedenheit aller Beteiligten bei.

Für das Wintersemester 2014/15 haben wir die Planung für die größten 270 Klausuren an der TU Berlin durchgeführt. Vorher erfolgte die Planung dezentral durch Teilabstimmungen der Anbieter. An Hochschulen derselben Größenordnung ist dafür eine Person Vollzeit beschäftigt, die sich mit vielen Personen abstimmen muss. Schätzungsweise haben wir dabei mindestens drei Personenjahre pro Jahr eingespart. Zusätzlich konnte eine erhebliche Qualitätssteigerung erzielt werden: Früher gab es eine bedeutende Anzahl an Klausuren, bei denen Studierende eines Studienganges gezwungen waren, zwei Klausuren am selben/folgenden Tag zu schreiben. Mit unserer Methode konnten wir solche Überschneidungen verhindern. Aktuell lösen wir die Klausurterminplanung mit 930 Klausuren an der TU München.

6.2 Lehrveranstaltungsplanung

Die RWTH Aachen und die TU Berlin erstellen ihre komplette Stunden- und Raumplanung während der Vorlesungszeit mit dem von uns entwickelten System. Anfangs gab es Widerstand gegen das neue Ressourcen-Management, inzwischen ist das neue Verfahren jedoch an beiden Universitäten fest etabliert. Eine Umfrage unter den Dozierenden der TU Berlin ergab, dass mehr als 85 % von ihnen mit der neuen Planung zufrieden sind. Da bei den alten Verfahren viele Dozierende mit ihren Terminen unzufrieden waren und die Umfrage direkt nach der Einführung durchgeführt wurde, ist das unseres Erachtens ein sehr gutes Ergebnis. Wir interpretieren das Umfrageergebnis als ein deutliches Indiz dafür, dass unsere Methode geeignet ist, konfliktfreie Stundenpläne für Studierende zu erstellen und Raumressourcen effizienter zu nutzen; außerdem zeigt es, dass so die Wünsche der Dozierenden erfüllen werden können und somit eine breite Akzeptanz an der Hochschule möglich ist.

An der RWTH Aaachen konnte durch unser System nicht nur der doppelte Abiturjahrgang trotz eines nicht fertiggestellten Hörsaal-Neubaus bewältigt werden, es konnten sogar mehr Studierende aufgenommen werden, als nach den alten Planungen vorgesehen war. Weiter erreichten die RWTH Aachen und die TU Berlin einen sehr viel effizienteren Einsatz der Raumressourcen – von den Veranstaltern seit langem „gebunkerte“ Räume wurden zurückgegeben und die zugewiesenen Räume an die tatsächlichen Teilnehmerzahlen angepasst. Die RWTH Aachen konnte teure Anmietungen bei steigenden Studierendenzahlen einsparen, die TU Berlin zieht das Abmieten teurer Flächen bei konstanter Studierendenzahl in Betracht. Die Kostenersparnis bei beiden Universitäten durch den effizienteren Raumeinsatz betrug ein Vielfaches der Projektkosten zur Einführung der neuen Stundenplanungsverfahren.

6.3 Tutorieneinteilung

Um 100 Studierende auf Tutorien überschneidungsfrei mit herkömmlichen Methoden zu verteilen, wird schätzungsweise ein Arbeitstag aufgewendet. Darin enthalten sind Organisation von Räumen, Verteilen der Tutoren/innen, Einteilung der Studierenden, Umverteilungen wegen späterer Überschneidungen und Erstellen der Teilnehmerlisten. Im WS 2012/13 haben wir an der TU Berlin mit 25.000 Tutorienplätze die bisher größte Einteilung vorgenommen. Das entspricht demnach ca. 1 Personenjahr. Durch unser Verfahren entfiel ein Großteil dieser Arbeiten. Selbst bei sehr konservativer Schätzung haben wir daher mindestens 1/2 Personenjahr eingespart. Pro Jahr ließe sich damit mindestens 1 Personenjahr sparen.

Noch stärker haben die Studierenden profitiert. Ihr Aufwand hat sich auf wenige Minuten reduziert, um sich zu allen Tutorien über eine Webmaske anzumelden. Bei der herkömmlichen Anmeldung in der Vorlesung oder der Eintragung in Listen schätzen wir den Zeitaufwand für die Studierenden auf 1 Stunde je Tutorienplatz.

Ein schwer quantifizierbares, aber wichtiges Qualitätsmerkmal unseres Verfahrens ist die gesteigerte Zufriedenheit der Studierenden und der Lehrenden. Die Studierenden erhalten einen überschneidungsfreien Stundenplan, der ihre Zeitpräferenzen bestmöglich berücksichtigt. Wartezeiten, Sorge um einen Platz und Ärger wegen Verschiebungen entfallen. Die Lehrenden werden von unnötigem Verwaltungsaufwand befreit.

6.4 Tutoreneinsatzplanung

Einige Jahre nach der Einführung der automatisierten optimierten Tutorieneinteilung an der TU Berlin kam bei den Anbietern großer Service-Veranstaltungen der Wunsch auf, die Tutoren-Einsatzplanung stärker an die Tutoreneinteilung zu koppeln. Bei schwankenden Studierendenzahlen sollten die Tutoren/innen flexibel Veranstaltungen zugewiesen werden können, um so jedes Semester die Tutorenkapazität den wirklichen Studierendenzahlen anpassen zu können. Es zeigte sich, dass man so nicht nur die Lehrkapazitäten gerechter auf die Veranstaltungen verteilen kann, sondern dass sogar ca. 10 % weniger Tutoren/innen benötigt wurden, um eine ausreichende Ausstattung der Veranstaltungen zu gewährleisten. Die Qualität der Lehre war dabei insgesamt höher: Wenig sinnvolle Überkapazität an Lehrpersonal konnte verhindert werden, und die Veranstaltungen, die sonst zuwenig Lehrpersonal gehabt hätten, konnten angemessen ausgestattet werden.

7 Ausblick

Der Schwerpunkt unserer Arbeit wird in der nahen Zukunft auf zwei Gebieten liegen: Wir sind dabei, einen umfassenden Arbeitsprozess in unser IT-System zu implementieren, der die Lehrveranstaltungsplanung, Tutorieneinteilung und Tutoren-Einsatzplanung in einem Gesamt-Arbeitprozess zusammenfasst. Entscheidend dabei ist es, den Arbeitsprozess so zu

entwickeln, dass kurzfristig auftretende Veränderungen der Rahmenbedingungen ressourceneffizienter und anwenderfreundlicher verarbeitet werden können. Der zweite Schwerpunkt wird eine Weiterentwicklung des Verteilungsalgorithmus zur Lehrveranstaltungsplanung sein. Viele Veranstalter haben ein großes Interesse an einer semesterübergreifenden Stabilität der Lehrveranstaltungspläne.

Literaturverzeichnis

[AA10] Alt, R.; Auth, G.: Campus Management System. Business & Information Systems Engineering, 2(3):187–190, 2010.

[BK09] Bick, M; Kummer, T: Referenzmodell zur Evaluation von Informationssystemen für ein integriertes Campus-Management. Hochschulmanagement, 4(4):108–114, 2009.

[BR06] Burke, E. K.; Rudová, H., Hrsg. PATAT 2006 – Proceedings of The 6th International Conference on the Practice and Theory of Automated Timetabling. Masaryk University, Brno, Czech Republic, Springer, 2006.

[Bu04] Burke, E. K., Hrsg. Proceedings of the 5th International Conference on the Practice and Theory of Automated Timetabling, Pittsburgh, Pa., 2004. Springer.

[CK96] Cooper, T. B.; Kingston, J. H.: The Complexity of Timetable Construction Problems. In (Burke, E. K.; Ross, P., Hrsg.): PATAT. Jgg. 1153 in Lecture Notes in Computer Science. Springer, S. 283–295, 1996.

[FV] Friedmann, J.; Verbeet, M.: , Blindflug ins Chaos. `http://www.spiegel.de/spiegel/0,1518,570180,00.html`.

[GJ78] Garey, M. R.; Johnson, D. S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman, Berlin, 1978.

[Go10] Gora, W.; Jeschke, S.; Lach, G.; Lübbe, J.; Pfeiffer, O.; Zorn, E.: Management and optimal distribution of large student numbers. In: Proceedings of the Education Engineering 2010. S. 1891–1896, 2010.

[La07] Lach, G.: Modelle und Algorithmen zur Optimierung der Raumvergabe der TU Berlin. Masterarbeit, Technische Universität Berlin, Institut für Mathematik, 2007.

[LL08] Lach, G.; Lübbecke, M.E.: Optimal University Course Timetables and the Partial Transversal Polytope. In (McGeoch, C.C., Hrsg.): 7th International Workshop on Efficient and Experimental Algorithms (WEA08). Jgg. 5038 in LNCS, Springer, Berlin, S. 235–248, 2008.

[Mc10] McCollum, B.; Schaerf, A.; Paechter, B.; McMullan, P.; Lewis, R.; Parkes, A. J.; Gaspero, L. Di; Qu, R.; Burke, E. K.: Setting the Research Agenda in Automated Timetabling: The Second International Timetabling Competition. INFORMS Journal on Computing, 22(1):120–130, 2010.

[MR14] Müller, T.; Rudov, H.: Real-life curriculum-based timetabling with elective courses and course sections. Annals of Operations Research, S. 1–18, jun 2014.

[Op] Opper, D.: , Studienstart bizarr. Bedingt einsatzbereit. `http://www.spiegel.de/unispiegel/studium/0,1518,449050,00.html`.

[ot] otr/dpa: , Hörsaal-Drängelei and Bafög-Stau. `http://www.spiegel.de/unispiegel/studium/0,1518,799551,00.html`.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

campus4uulm – Flexibles und integriertes Campus-Management mit Standardsoftware?

Stefan Wesner[1], Claudia Pauli[2], Jens Kohlmeyer[3], Beate Quester-Brüning[4] und Marcus Link[5]

Abstract: Dieser Artikel gibt einen Einblick in die ersten Ergebnisse des Projektes campus4uulm. In diesem Projekt wird das Ziel verfolgt, nicht nur die bestehende Campus-Management-Lösung zu ersetzen, sondern im Zuge der Einführung dieser neuen Software auch die über die Jahre etablierten Geschäftsprozesse kritisch zu reflektieren. Das Projekt wird nicht als reines IT-Projekt sondern insbesondere als Organisationsentwicklungs- und Prozessoptimierungsprojekt verstanden. In diesem Beitrag wird, ausgehend von einer für das Projekt festgelegten, über das übliche Verständnis hinausgehenden Definition des integrierten Campus-Managements, die vorbereitenden Schritte, das Umsetzungskonzept bis hin zum gewählten iterativen Vorgehensmodell und den Implementierungsansätzen für das sogenannte integrative Campus-Management-System vorgestellt.

Keywords: Campus-Management, Student Lifecycle Management, Prozessoptimierung, Organisationsentwicklung, Harmonisierung der IT-Systemlandschaft

1 Einleitung

An vielen Hochschulen und Universitäten laufen derzeit Projekte zur Vorbereitung oder Implementierung neuer Campus-Management-Lösungen. Beschäftigt man sich mit dieser Fragestellung wird schnell offensichtlich, dass die Einführung eines solchen komplexen, alle Bereiche der akademischen Struktur berührenden IT-Systems nicht ohne die Bereitschaft, dies auch als Organisationsentwicklungsprojekt aufzufassen, gelingen kann (siehe dazu auch [Ja09, De09]). Dies bedingt zunächst einen erhöhten Mehraufwand für alle Beteiligten der Universität während der Projektlaufzeit, ist aber für die Nachhaltigkeit der Organisationsentwicklung unerlässlich.

Neben den zentral bereitgestellten Komponenten gibt es auch eine Vielzahl spezieller Systeme, die individuelle Abläufe und Prozesse an den Fakultäten, Instituten und anderen an der Lehre beteiligten Einrichtungen unterstützen. Diese Systeme sind oft über Schnittstellen angebunden, die wenn überhaupt nur eine halb-automatisierte Datenübertragung erlau-

[1] Universität Ulm, Kommunikations- und Informationszentrum (kiz), Albert-Einstein-Allee 11, 89081 Ulm, stefan.wesner@uni-ulm.de
[2] Universität Ulm, Kommunikations- und Informationszentrum (kiz), Albert-Einstein-Allee 11, 89081 Ulm, claudia.pauli@uni-ulm.de
[3] Universität Ulm, Service Gruppe Informatik (SGI), James-Franck-Ring, 89081 Ulm, jens.kohlmeyer@uni-ulm.de
[4] Universität Ulm, Kommunikations- und Informationszentrum (kiz), Albert-Einstein-Allee 11, 89081 Ulm, beate.quester-bruening@uni-ulm.de
[5] Universität Ulm, Kommunikations- und Informationszentrum (kiz), Albert-Einstein-Allee 11, 89081 Ulm, marcus.link@uni-ulm.de

ben. Diese Anbindung ist darüber hinaus oft nur uni-direktional vom Campus-Management-System zum Drittsystem oder umgekehrt realisierbar.

Das Projekt campus4uulm an der Universität Ulm hat sich nicht nur zum Ziel gesetzt, das aus Einzelmodulen bestehende bisherige Campus-Management in eine integrierte Standardlösung zu überführen, sondern im Zuge dieser Umstellung auch eine weiterführende Vereinheitlichung und Integration auf Prozessebene zu erreichen.

Ausgehend von einer für dieses Projekt gewählten Definition des *integrativen* Campus-Managements wird in diesem Beitrag im direkt folgenden Kapitel dargestellt, welche Zielsetzungen bei der Neuausrichtung des Campus-Managements adressiert werden und durch welche Ansätze der Organisationsentwicklung Rechnung getragen wird. Darüber hinaus werden die daraus resultierenden Implikationen für die IT-Systemlandschaft dargestellt.

Im anschließenden Kapitel 3 wird die daraus abgeleitete grobe Systemarchitektur und das Umsetzungskonzept inklusive der verschiedenen Projektphasen dargestellt. In diesem Kapitel wird auch darauf eingegangen, wie sich dieses abstrakte Konzept mit den gewählten IT-Komponenten und -Produkten der beteiligten Anbieter spezifisch ausprägt. Im Anschluss wird der gewählte iterative Ansatz dargestellt und insbesondere die Rolle der Key User und Prozessverantwortlichen erläutert.

Zum Abschluss werden erste Erfahrungen bei der Umsetzung des Projektes dargestellt.

2 Integratives Campus-Management

Der Begriff integriertes Campus-Management wird oft gleichgesetzt mit dem Übergang einer aus Einzelmodulen bestehenden Lösung für die verschiedenen Aufgaben im administrativen Bereich von Studium & Lehre, wie z.B. Bewerbung & Zulassung, Prüfungsorganisation oder Lehrveranstaltungsmanagement, zu einem homogenen System. Für das Projekt campus4uulm wurde der Begriff des integrierten Campus-Managements bewusst weiter gefasst hin zu einem sogenannten integrativen Campus-Management. Nach unserem Verständnis kann die Integration nicht bei den bisher üblichen Funktionsblöcken, wie z.B. den HIS-Altsystemen, aufhören, sondern muss in einer offenen Architektur auch darüber hinaus gehende Integrationsmöglichkeiten bieten. Neben der Anbindung an Finanzdaten und Personalstammdaten, z.B. auch externer Einrichtungen bei Kooperationsstudiengängen oder regionalen Kooperationsverbünden, sind insbesondere Systeme, die zur weiteren Unterstützung der Lehraufgaben verwendet werden, zu integrieren. Neben größeren zentralen Systemen, zu denen eine generell nutzbare Schnittstelle realisiert wird, wie zu einem Learning-Management-System, z.B. Moodle[DT03, Mo], oder zum Web-Content-Management-System für den Webauftritt der Universität, gibt es auch auf Fakultäts-, Instituts- oder Mitarbeiterebene eine breite Palette an kleineren und größeren Werkzeugen zur Ergänzung der bestehenden Funktionalitäten (siehe Abbildung 1). Sinnvollerweise sollte der Nutzerkreis derartiger Werkzeuge von Beginn an in das Projekt eingebunden werden.

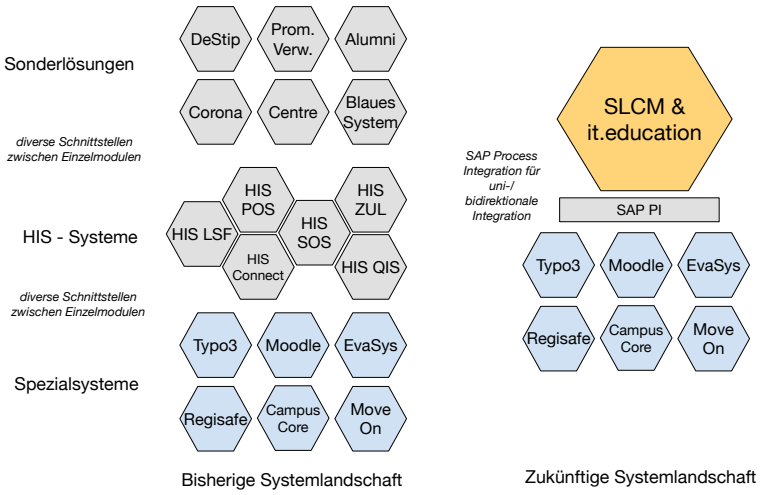


Abb. 1: Grobkonzept Harmonisierung der IT-Systemlandschaft im Campus-Management

Ein integratives Campus-Management muss daher nicht nur die Funktionalitäten bestehender Systeme ersetzen, sondern muss auch durch geeignete Schnittstellen andere komplexe Systeme so einbinden, dass diese nicht nur als Konsumenten von Daten oder für den Bulk-Import verwendet werden können. Wir folgen daher in unserer Definition eines integrativen Campus-Management mehr der Definition von Integrierten Management Systemen (IMS) [Ho13]. Dabei ist die Zielsetzung, alle IT-Systeme innerhalb der Universität im administrativen Bereich von Studium & Lehre so zu integrieren, dass Geschäftsprozesse durch die verschiedenen Akteure nahtlos durchgeführt werden können, ohne Aktivitäten zu duplizieren oder Daten mehrfach erfassen zu müssen. Das Zielsystem soll also zur Harmonisierung der Systemlandschaft im IT-Bereich beitragen und gleichzeitig auch dem Anwender ermöglichen, für einen Vorgang so wenig verschiedene IT-Systeme wie möglich einsetzen zu müssen.

2.1 Harmonisierung der Prozesse

Der oben genannte Anspruch an die Integration erforderte im Vorfeld der Auswahl eines Software-Anbieters einiges an Vorarbeiten. So war die Erfassung bestehender externer Werkzeuge und Systeme notwendig, die in irgendeiner Form als Datenkonsumenten oder -lieferanten für das bisherige Campus-Management-System dienen. Die Orientierung an den Geschäftsprozessen setzte nicht nur deren Analyse, Erfassung und Dokumentation voraus, sondern machte auch eine kritische Betrachtung und Bewertung von derzeit unterschiedlich ablaufenden Prozessen für gleiche oder ähnliche Vorgänge nötig. Als Beispiel hierfür seien die Kursanmeldungs- und Kursverwaltungsverfahren in der Medizinischen Fakultät und im Sprachenzentrum des Departments für Philosophie, Sprachen, Geisteswissenschaften und allgemeine Weiterbildung genannt. Durch die kritische Hinterfragung der dort bestehenden, sehr unterschiedlich gehandhabten Prozesse konnte eine gemein-

same vereinheitlichte Vorgehensweise gefunden werden, die im Rahmen des Projektes IT-technisch umgesetzt wird und den Betrieb und die Pflege der eigenen Speziallösungen künftig obsolet werden lässt.

Zur Erfassung und Dokumentation der Ergebnisse in einem Prozesshandbuch wurden neben einer Expertengruppe mit IT-Hintergrund und Erfahrungen bei der Prozessmodellierung – vgl. Abbildung 2 für einen Beispielprozess – insbesondere auch Fachleute für die verschiedenen universitären Aufgabenbereiche, so genannte *Key User*, identifiziert und eingebunden. Ein solches Vorgehensmodell ist bei der Einführung von komplexen Softwaresystemen ein üblicher Ansatz und wird für das Campus-Management unter anderem auch in [SAA12] vorgeschlagen. Im Projekt campus4uulm wurden im Rahmen der Organisationsentwicklung zahlreiche Gespräche und Workshops mit den Key Usern und Experten durchgeführt, die Informationen gesammelt, modelliert und nach Verfeinerung auch entsprechend abgenommen.

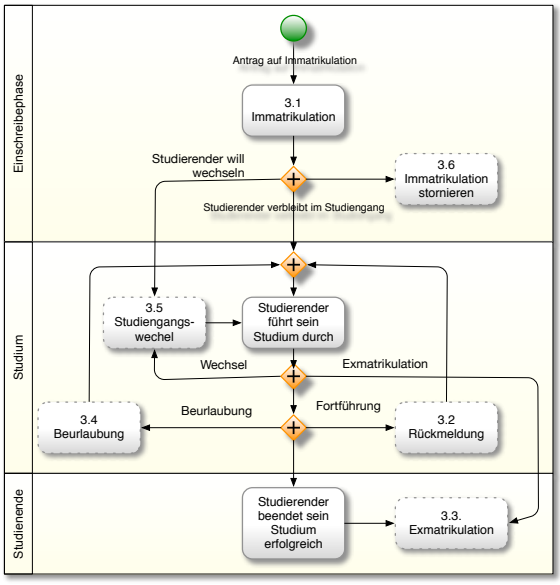


Abb. 2: Beispiel für einem modellierten Prozess mit Unterprozessen

Letzteres zeigt einen weiteren Aspekt, der bei der Vorgehensweise unter *integrativ* verstanden wird: Die Nutzergruppen des Systems haben sehr unterschiedliche Anforderungen und Prioritäten an ein künftiges System, von denen nicht alle im verfügbaren Zeit- und Kostenrahmen realisiert werden können. Es ist daher essentiell für eine vollständige Erfassung der Anforderungen und auch für die Akzeptanz des Nachfolgesystems, alle Anwender möglichst frühzeitig in den Migrationsprozess zu integrieren. Dazu wurden die aus den einzelnen Anwendergruppen stammenden Key User als repräsentative Vertreter (“stakeholder“) in den Prozess eingebunden. Beim Umsetzungskonzept wurde deren Rolle durch die Etablierung von sogenannten Prozessverantwortlichen (siehe Abschnitt 3.3) weiter gestärkt.

2.2 Etablierung eines führenden zentralen Systems

In [Ra09] wird die These aufgestellt, dass bei der Einführung eines Integrierten Campus-Managements grundsätzlich die Optionen eines Gesamtsystems oder der Integration verschiedener Standardsoftware-Komponenten besteht. Im Projekt campus4uulm wird ein Ansatz gewählt, der sich keinem dieser beiden Extreme zuordnen lässt. Die in [Ra09] auch als wichtig angesehene Definition eines *führenden Systems* bei der Datenhaltung ist auch ein essentieller Grundsatz des gewählten Lösungsansatzes. Allerdings sehen wir das neue System nicht als eine Integrationslösung von Einzelkomponenten, sondern verfolgen den Ansatz eines dominierenden zentralen Systems für die Unterstützung der administrativen Kernprozessen im Campus-Management, wobei spezielle Aufgaben teilweise an externe Systeme ausgelagert werden können. Bei der Integration externer Systeme ist grundsätzlich festzulegen, ob diese Daten lediglich konsumieren oder eine komplexe Interaktion über die Schnittstelle mit dem Campus-Management-System zu realisieren ist. Die beim bisherigen Campus-Management-System realisierten Anbindungen auf Basis von Textdateien soll, soweit möglich, als automatisierte Kopplung der Systeme in Form einer Programmierschnittstelle bereitgestellt werden. Entscheidend hierbei ist, dass es – neben den untergeordneten Datenkonsumenten, denen Daten (nur) zur externen Nutzung zur Verfügung gestellt werden (z.B. Liste angemeldeter Studierender zur Weiternutzung in Excel), oder Datenquellen, die Daten für einen einmaligen oder regelmäßigen Import zur Verfügung stellen (z.B. Raumlisten) – weitere, in das Campus-Management-System eng integrierte Systeme gibt. Kann die Kopplung nicht derart realisiert werden, kann diese Lücke nur durch nicht zu unterschätzenden, manuellen Aufwand geschlossen werden, wodurch Medienbrüche und die damit einhergehende Gefahr von Dateninkonsistenzen entstehen.

Es werden zwischen den am Campus-Management beteiligten Systemen programmierbare Schnittstellen benötigt, die es erlauben, einen ggf. auch bidirektionalen Datenaustausch untereinander zu realisieren. Ob dieser komplett synchron, z.B. durch die Nutzung von REST-basierten Schnittstellen, durch den direkten Zugriff auf Datenbankebene oder mit Hilfe von Asynchronous Message Queuing Protocol (AMQP) basierten Systemen realisiert werden, spielt dabei für die Architektur keine entscheidende Rolle.

3 Umsetzungskonzept

Das grundlegende Element des Umsetzungskonzeptes ist, wie oben ausgeführt, die Realisierung eines offenen "integrativen" Systems. Alle Systeme im Bereich des Campus-Managements auf dem Markt sehen für bestimmte Prozesse Standardabläufe vor. Unter einer konfigurierbaren und offenen Systemarchitektur wird nun verstanden, dass diese parametrierbare Referenzprozesse und Grundfunktionalitäten mitbringt. Dazu müssen von der Datenbankebene bis hin zu den konkret ausgeprägten Prozessabläufen Anpassungen an die Bedürfnisse der Hochschule möglich sein, ohne dass eine individuelle Softwareentwicklung vorgenommen werden muss.

Da sich allerdings nicht alle notwendigen Funktionalitäten nach dem oben angeführten Modell der Konfiguration realisieren lassen, ist es nicht zu vermeiden, entweder Eigenentwicklungen bzw. Erweiterungen auf Code-Ebene am Campus-Management-System vorzunehmen oder diese speziellen Funktionalitäten durch darauf spezialisierte Systeme abzubilden. Jede individuelle, nur für eine Hochschule durchgeführte Anpassung bringt jedoch einen schwer kalkulierbaren Wartungsaufwand mit sich und führt damit zu erhöhten Wartungskosten. Daher muss auch im Rahmen der Organisationsentwicklung an der Hochschule ein Mindestmaß an Bereitschaft für die Einführung von Standardprozessen vorhanden sein.

Aus dem Blickwinkel der Sicherung der Nachhaltigkeit ist bei der Realisierung umfangreicher spezieller Funktionalitäten, die nicht durch Standardfunktionalitäten im führenden zentralen Campus-Management-System, abgedeckt sind, weitestgehend eine Anbindung von externen Systemen zu forcieren. Das setzt beim Campus-Management-System die Möglichkeit einer flexiblen Schnittstellen-Implementierung basierend auf offenen Standards voraus. Diese Vorgehensweise erfordert lediglich, die bereits oben beschriebenen Schnittstellen zum Datenaustausch zu realisieren und diese zu warten. Als Beispiel hierfür kann ein eLearning-System wie Moodle dienen, dessen umfangreiche Funktionen des elektronisch unterstützten Lernens nicht in ein Campus-Management-System gehören, dessen enge Anbindung aber wünschenswert ist. Ein Beispielszenario für eine solche Anbindung kann sein, dass eine neue Veranstaltung im Campus-Management-System automatisch und zeitnah auch im Moodle-System verfügbar ist, die definierten Verantwortlichkeiten dafür automatisch auch dort erscheinen, aber auch umgekehrt erfolgreich durchgeführte Tests als Vorleistung ins Campus-Management übernommen werden können.

3.1 Grobe Systemarchitektur

Die grobe Systemarchitektur geht von einer üblichen mehrschichtigen Architektur mit einer Präsentationsschicht, einer Anwendungslogikschicht und einer Persistenzschicht aus.

In der Präsentationsschicht werden die Daten und Eingabemöglichkeiten für die verschiedenen Nutzergruppen je nach eingesetztem Gerät geeignet präsentiert. Hierbei ist nicht für jede Nutzergruppe die ganze Bandbreite an Endgeräten zu unterstützen (z.B. sind für die administrativen Backend-Aufgaben in der Verwaltung mobile Geräte eher nicht im Einsatz). Für den Nutzerkreis der Studierenden allerdings ist z.B. für die Abfrage von Noten oder für die aktive Benachrichtigungen bei Verfügbarkeit von Prüfungsergebnissen die Bereitstellung einer Smartphone-Schnittstelle sicher wünschenswert.

Neben der direkt für den Endanwender gedachten Präsentationsschicht ist für das Berichtswesen und die Kopplung mit systeminternen Komponenten eine bereits oben genannte Programmier- und/oder Datenschnittstelle notwendig. Die Anwendungslogikschicht stellt sowohl für die Geschäftsprozesse als auch für die Bedienung von Schnittstellen definierte Funktionen für den kontrollierten Zugriff auf die Daten der Persistenzschicht zur Verfügung. Ein direkter Zugriff auf die ganz unten liegende Persistenzschicht in Form einer Datenbank ist nicht erlaubt. Ein lesender Zugriff externer Komponenten wird ebenfalls über die Anwendungslogikschicht geregelt.

3.2 Produktspezifische Systemarchitektur

Ausgehend von dieser grundlegenden mehrschichtigen Architektur wurde nun die an der Universität Ulm bereits produktive SAP ERP Lösung um das SAP SLcM [SA] Modul erweitert. Die Branchenlösung it.education [it] ist als eine Zusatzimplementierung und als Referenzkonfiguration des SAP SLcM Moduls anzusehen: Ihre Zusatzfunktionalität minimiert den Zeit- und Ressourcenaufwand bezüglich der sich aus den individuellen Anforderungen ergebenden kundenspezifischen Implementierungen. Ebenso verringert die it.education Referenzkonfiguration den Aufwand bezüglich der durchzuführenden individuellen Konfigurationseinstellungen für das kundenspezifische System.

In Abbildung 3 ist dargestellt, dass das SAP SLcM Modul inklusive der it.education Erweiterungen auf gleicher Ebene wie eine Vielzahl von anderen bereits verfügbaren SAP Modulen eingeordnet wird. Die Kommunikation zwischen diesen Komponenten kann durch Nutzung SAP-interner Funktionalität realisiert werden. Da alle Module auf einer gemeinsamen Datenbank arbeiten, ist im Prinzip kein expliziter Datenaustausch zwischen den Modulen nötig. Für das Berichtswesen im SAP BW existieren Standardschnittstellen zur Datenextraktion und -überführung, die auf gleiche Weise sowohl im Campus-Management als auch im Finanz- und Controllingbereich genutzt werden können.

Ebenso können Schnittstellen zu Non-SAP Systemen auf standardisierte Art und Weise – z.B. über das SAP Process Integration Management Tool (PI) – realisiert werden; dabei ist es möglich, sowohl asynchrone als auch synchrone Datenaustausche zu realisieren, Daten in Echtzeit oder zeitlich getriggert auszutauschen sowie direkt Funktionen im SAP ERP und – je nach deren Möglichkeiten – in Non-SAP Systemen anzusprechen und somit systemübergreifende Geschäftsprozesse zu implementieren, ohne dass sich für den Benutzer ein Systembruch ergibt.

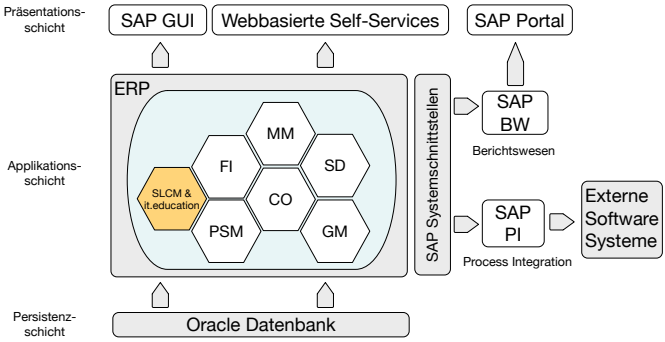


Abb. 3: Ansatz zur Harmonisierung der IT Systemlandschaft

Kernelement dieses Lösungsansatzes ist eine weitestgehende Vereinheitlichung der IT-Systemlandschaft auf Basis von SAP-Modulen und der damit einhergehenden Zusammenführung der bisherigen Gruppen für den Betrieb und die Anwendungs- und Nutzerbetreuung bzgl. der SAP- und der HIS-Systeme in ein gemeinsames Kompetenzzentrum für diese Dienste. Darüber hinaus wurde für eine möglichst schnelle Umsetzung der bisher durch die HIS-Systeme bereitgestellten Funktionalitäten auf das it.education-Framework

gesetzt, das oberhalb der bereits im SLcM-Modul enthaltenen Funktionen bereits vorgefertigte und konfigurierbare Prozesse, aber auch zusätzliche Unterstützung im Bereich der Benutzerschnittstelle und Anbindung externer Systeme mit sich bringt. Ein weiterer Vorteil ist, dass sich die Schnittstellenrealisierungen und deren Pflege auf SAP-interne Schnittstellen zu verschiedenen Modulen und die Anbindung an it.education über SAP PI reduzieren. Durch diese erhebliche Konsolidierung der Schnittstellen zum Campus-Management-System wird erwartet, dass sich deren Wartung vereinfacht und gleichzeitig in der Anbindung eine bessere und erweiterte Funktionalität realisiert werden kann.

Die Integration der einzelnen Komponenten in einer konkreten Installationssicht konnte dabei zu einem hohen Maß ohne Veränderungen von der bisherigen Struktur der bereits seit längerem betriebenen SAP-Modulen übernommen werden. Hier ist insbesondere die Schwierigkeit zu adressieren, dass sich – neben der bisher vorherrschenden Nutzung durch interne Nutzer aus dem Verwaltungsnetz und dem Universitätsnetz – durch die Nutzung durch Studierende nicht nur die Nutzerzahl drastisch erhöhen wird, sondern auch der Systemzugriff aus Fremdnetzen oder auch extern zugänglichen Netzen wie dem Eduroam-Netz des DFN-Vereins erfolgen wird. Dies erfordert somit nicht nur eine Erhöhung der verfügbaren Ressourcen am Frontend für die gestiegenen Benutzerzahlen und eine entsprechenden Anpassung der Backend-Infrastruktur sondern auch die Überprüfung des bisherigen Sicherheitskonzeptes.

3.3 Anforderungs- und Abnahmemanagement

Wie dargestellt wird bei der Umsetzung des Projektes sehr stark auf den Einsatz der Standardsoftware SAP SLcM und auf die bereits mit Grundfunktionen und Referenzprozessen ausgestattete Branchenlösung it.education gesetzt. Grundlage für die Realisierung dieses Ansatzes ist, die notwendigen Prozesse und Abläufe mit denjenigen des Systems abzugleichen, um eine Einschätzung zu ermöglichen, in wie weit diese von den bestehenden verschiedenen Softwarekomponenten bereits umgesetzt sind, welche davon mit Hilfe welcher Komponenten oder externer Systeme umgesetzt werden sollen und an welcher Stelle Neuentwicklungen oder Schnittstellenimplementierungen notwendig sind.

Ausgangspunkt für das Projekt bildeten daher wie in Abschnitt 2.1 ausgeführt die mit den Key Usern analysierten Ist-Prozesse und die Dokumentation der Soll-Prozesse in einem umfangreichen Prozesshandbuch. Diese systemunabhängige Darstellung musste nun nach Auswahl eines Produktanbieters auf die mit dem neuen System möglichen Prozesse und Funktionen detailliert abgestimmt werden. Dazu wurden thematisch strukturierte Workshops mit dem Dienstleister konzipiert und abgehalten, die sich einerseits an den Prozessdomänen, wie etwa Bewerber- und Studierendenmanagement oder Lehrveranstaltungs- und Prüfungsmanagement, des Prozesshandbuches orientierten, andererseits aber auch schon die dafür notwendige Basis bei der systemtechnischen Einrichtung der Organisationsstruktur und der Akademischen Struktur für die jeweiligen Prozessdomänen mit berücksichtigten. Darüber hinaus wurden Themen zu Schnittstellen, z.B. zur Lehr- und Lernplattform Moodle, zum Web Content Management System Typo3 und zum eigenentwickelten Identitätsmanagementsystem inkl. Fragen zur Authentifizierung sowie zur

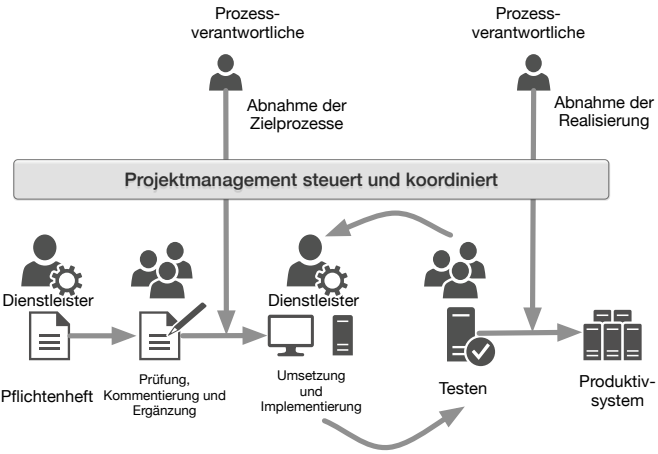


Abb. 4: Rollen und Aufgaben bei Entwicklung und Abnahme in Absprache mit dem Anbieter

Datenmigration in separaten Workshops adressiert. Dabei hat es sich als äußerst hilfreich und effizient herausgestellt, zu diesen Workshops in zusätzlichen Vorab-Treffen und Vorbereitungsgesprächen zwischen den technischen Experten des Projektmanagementteams und den jeweiligen Key Usern im Vorfeld die Prozesse noch einmal zu diskutieren und die besonders wichtigen Aspekte und Kriterien sowie die Standardisierungspotenziale herauszuarbeiten. Im Anschluss fanden dann jeweils die Workshops zusammen mit dem Dienstleister statt, in denen oftmals durch interaktive Elemente direkt an der Softwarelösung die Realisierungsmöglichkeiten aufgezeigt wurden. Die Ergebnisse der einzelnen Workshops wurden sorgfältig in so genannten Design-Dokumenten festgehalten, die durch die Key User und Experten geprüft und häufig in zusätzlichen so genannten "Rücksprache-Workshops" ein weiteres Mal mit dem Dienstleister abgestimmt wurden. Nach diesem sehr aufwändigen und intensiven Vorbereitungsprozess für das campus4uulm-Projekt wurden diese Design-Dokumente schließlich in ein so genanntes Pflichtenheft eingearbeitet, das das Ergebnis des als Initialisierungsphase bezeichneten Projektabschnitts darstellt. Dieses Pflichtenheft wurde mittlerweile von den Key Usern und einer Expertengruppe geprüft und den für jeden Prozess identifizierten Prozessverantwortlichen zur Abnahme vorgelegt. Wie aus Abbildung 4 ersichtlich wird, stellt dieses Vorgehen generell den formalen Abnahmeschritt auf Konzeptebene dar, bevor für die einzelnen Prozesse eine Realisierungsphase begonnen wird.

Im Sinne eines iterativen Vorgehensmodells stellt dieser Meilenstein den Übergang von der Initialisierungsphase in die Umsetzungsphasen dar (siehe dazu auch Tabelle 1). Dabei wird in einem ersten Iterationsschritt (Implementierungsphase 1) mit dem thematischen Schwerpunkt "Bereitstellung von Modulbeschreibungen" bewusst ein nicht an den Semesterrhytmus gebundenes Thema gewählt, um dem Aufbau der notwendigen Infrastruktur (Bereitstellung von Entwicklungs-, Test- und Produktionsservern), der Bereitstellung der zusätzlichen Softwaremodule und notwendigen Systemerweiterungen sowie der Aktualisierung auf notwendige neuer Release-Versionen der Basissoftware genügend Zeit

einräumen zu können. Im Rahmen dieser ersten Implementierungsphase wird insbesondere auch untersucht, ob alle SAP-Komponenten auf einem System zusammengeführt und betrieben werden sollen oder ob aufgrund der sehr unterschiedlichen Nutzergruppen der SLcM-Komponenten und der anderen ERP-Komponenten ein Parallelbetrieb nicht die bessere Alternative ist. Dabei soll der offensichtliche Vorteil des halbierten Aufwands und der halbierten Kosten für den Betrieb von einem statt zwei SAP ERP-Basissystemen und deren darunter liegenden Serversystemen gegen den wichtigsten Vorteil einer Entkopplung der Releasezyklen, die im Bereich von SLcM deutlich kürzer ausfallen als bei den für das Finanzwesen wichtigen Modulen, abgewogen werden. Darüber hinaus wird in diesem Schritt auch mit dem Aufbau der entsprechenden unterstützenden Softwareinfrastruktur das Configurations-Management und die Kommunikations- und Projektfortschrittsüberwachung auf Softwareebene etabliert. Die primäre Zielsetzung besteht darin, erste Funktionalitäten im neuen System bereitzustellen und bei der Umsetzung nicht nur die Basisinfrastruktur zu realisieren, sondern in gleichem Maße auch Erfahrungen bei der Realisierung zu erlangen, um eine bessere Grundlage für die Iterations- und Zeitplanung zu erhalten. Diese Phase kann als "Inception/Elaboration Phase" angesehen werden. Die

Phase	Einordnung	Ziele und Funktionen
Vorbereitungsphase	Inception	Modellierung der Ist-Prozesse, Reflektion der Prozessabläufe und Festlegung von produktunabhängigen Soll Prozessen
Implementierungsphase1	Elaboration	Realisierung erster Funktionalitäten mit den Schwerpunkt auf Module- und Modulhandbücher. Parallel Aufbau der Basisinfrastruktur und Schulung des technischen Personals
Implementierungsphase 2	Construction & Transition	Realisierung von einer breiteren Palette von Funktionen und Ablösung von Kernfunktionen durch das neue System. Dies umfasst u.a. Veranstaltungsmanagement, Studierendenverwaltung, Prüfungswesen
Implementierungsphase 3	Construction & Transition	Realisierung von Bewerbung & Zulassung inkl. DoSV Anbindung und Doktorandenmanagement

Tab. 1: Geplante Implementierungsphasen und -schwerpunkte

folgenden Schritte zeigen bei ihren Zielsetzungen dann einen klaren Fokus auf die Umsetzung und die Realisierung von Funktionen, die bereits zur Ablösung von bisherigen Funktionalitäten in den HIS-Altsystemen führen sollen. Dabei sind insbesondere die Datenmigration und der parallele Systembetrieb eine besondere Herausforderung. Daher sind in diesen Phasen die Funktionalitäten so gewählt, dass nach Möglichkeit mit Inbetriebnahme von Teilfunktionen eine Nutzergruppe komplett auf das neue System umsteigen kann und eine parallele Nutzung von zwei Systemen auf ein Minimum beschränkt wird.

Bevor die Definition von Abbildungsregeln zur Überführung von bisherigen Daten in die Datenbasis (siehe dazu auch Tabelle 2) des neuen Zielsystems begonnen werden kann, ist festzulegen, welche Daten überhaupt ins neue System übernommen werden sollen und bei welchen Datenbeständen eine manuelle Übernahme auch als Gelegenheit dienen kann, die Datenqualität zu verbessern.

Datenart	Migrationsansatz
Inaktive Daten	Unter diese Rubrik fallen alle Daten von Studierenden, Doktoranden, deren Mitgliedschaft an der Universität beendet ist. Eine Überführung dieser Daten in das neue System wird nicht durchgeführt. Diese Daten sind in den jeweiligen Studierenden-Akten ausreichen dokumentiert und archiviert.
Auslaufende Studiengänge	Studierende, die sich noch in bereits seit längerem abgelaufenen Studiengängen befinden, die vor der Bachelor/Master-Umstellung angeboten wurden (wie z.B. Diplom-Studiengänge), werden bis zum Studienabschluss im alten System weitergeführt.
Daten hoher Qualität	Für Daten, die bisher in Systemen gepflegt werden, die eine hohe Datenqualität sicherstellen, werden Abbildungsregeln definiert mit dem Ziel, diese Daten automatisiert zu übernehmen. Sollten zusätzliche Datenfelder im Zielsystem gepflegt werden, wird nach Möglichkeit versucht, aus den vorhandenen Daten diese Felder zu belegen bzw. mit Standardwerten hier vollständige Datensätze zu generieren.
Daten geringer Qualität	Daten, die mit Systemen erfasst wurden, die viele Freiheiten bei der Eingabe erlauben, bzw. Datenbeständen, die im Altsystem aus derartigen Quellen einbezogen wurden, sollen nach Möglichkeit bei der Migration bereinigt oder vereinheitlicht werden. Typische Vertreter dafür sind in Textdateien oder Spreadsheets erfasste Daten, z.B. von externen Doktoranden.

Tab. 2: Umgang mit Altdaten

4 Fazit und Ausblick

Die bisherigen Erfahrungen haben gezeigt, dass der mit einem Jahr Dauer sehr zeitaufwändige Vorbereitungsprozess, die bestehenden Abläufe zu analysieren und graphisch zu modellieren, aus zwei Gründen maßgeblich für den Erfolg des Projektes ist:

1. Es werden die Arbeitsgruppen und Personen identifiziert, die zum Start des Projekts bereits eine etablierte Basis für die Zusammenarbeit gefunden haben.
2. Häufig haben die Nutzer des bestehenden Systems trotz langjähriger Arbeit in den verschiedenen Bereichen keine Sicht auf den gesamten Geschäftsablauf. Es sind in der Regel nur die Personen/Funktionen bekannt, die Arbeitsergebnisse nutzen oder auf deren Vorarbeiten aufgebaut wird. Die graphische Darstellung des Gesamtprozesses erlaubt einen Blickwinkel aus Sicht aller Fakultäten und der Universität.

Im bisherigen Projektverlauf ist es darüber hinaus gelungen, die verschiedenen Anforderungen aus den unterschiedlichen Studiengängen und -ordnungen herauszuarbeiten und deren Umsetzung durch Konfigurationsmöglichkeiten des IT-Systems mit dem Dienstleister abzustimmen, um so die zu erwartenden Sonderentwicklungen auf einem niedrigen Niveau zu halten. Die Zielsetzung, die Anbindung von externen Systemen enger zu gestalten, ist für das eingesetzte Learning Management System (LMS) auf Basis von Moodle spezifiziert, und auch die automatisierte Anbindung von Inhalten aus dem Campus-Management-System im neuen Web-Auftritt der Universität ist vorgesehen.

Bisher kann der Ansatz eines "integrativen" Campus-Managements durch Einbindung der Key User und Prozessverantwortlichen bei der Spezifikation und Abnahme von Prozessen als erfolgreich bewertet werden. Auch auf der technischen Ebene konnte neben der bereits oben erwähnten engeren Kopplung mit externen Systemen erreicht werden, Abläufe, die bisher eher weniger gut durch IT-Lösungen unterstützt wurden, zu integrieren und deren Prozesse stärker zu vereinheitlichen. Dies beruht insbesondere auf der Tatsache, dass das campus4uulm-Projekt von vorne herein als ein Organisationsentwicklungsprojekt verstanden wurde. Die Notwendigkeit einer strukturierten Einführung unter Mitnahme aller Nutzergruppen wird auch in [De09] als Erfolgskriterium genannt. Die Einbeziehung aller Key User an der Hochschule bedingt allerdings einen hohen Organisationsaufwand, sowohl universitätsintern beim Projektmanagementteam als auch beim externen Dienstleister, damit die erforderlichen Personenkreise zu den jeweiligen Projektphasen mit ausreichend Zeit verfügbar sind. Dies erfordert einen hohen Zeitaufwand, der von allen universitätsinternen Beteiligten neben den originär übertragenen Aufgaben aufzubringen ist und erzeugt beim Dienstleister zusätzliche Kosten.

In der nun begonnenen Implementierungsphase treten neben der Akzeptanz durch die Endanwender insbesondere auch Fragen in den Vordergrund, ob die Annahmen zur Auslegung der Hardwareinfrastruktur ausreichende Ressourcen für eine effiziente und schnelle Bearbeitung ermöglichen. Auch wird sich noch zeigen müssen, ob sich der gewählte Ansatz einer engen Kopplung zwischen dem Campus-Management-System mit anderen zentralen Diensten (LMS, Web Auftritt, ...) neben den offenkundigen Vorteilen in der Praxis bewährt.

Aufgrund der engen Verzahnung des Campus-Managements mit allen universitären Bereichen und seiner starken Bedeutung hinsichtlich einer Organisationsentwicklung wurde dieses interne Projekt in seiner Struktur und Governance ähnlich einem externen Projekt dieser Größenordnung angelegt. Neben der Aufteilung der technischen Verantwortlichkeit in verschiedene Bereiche (Arbeitspakete), gibt es neben einem Projektmanagementteam, das die technische Steuerung und das "day-to-day"-Management verantwortet, noch den Fachbeirat als ultimatives Entscheidungsgremium für die Abnahme der Fachkonzepte und den Lenkungsausschuss, der neben Vertretern des Präsidiums und der Projektleitung auch Vertreter der beteiligten Firmen einbindet. Im Lenkungsausschuss wird die formale Abnahme von Meilensteinen beschlossen und gegebenenfalls eine Risikobewertung auf Projektebene durchgeführt. Parallel zu dieser Struktur findet in einem regelmäßigen Prozess eine Identifikation und Bewertung von Projektrisiken auf technischer Ebene statt, um deren Auswirkungen auf den weiteren Projektverlauf frühzeitig zu erkennen und zu beheben.

5 Acknowledgements

Die in diesem Beitrag vorgestellten Ergebnisse sind teilweise durch das vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg geförderte Projekt bw-CMS des Arbeitskreises IQF-Campusmanagement der Universitäten des Landes Baden-Württemberg zur Umsetzung der thematischen IQF-Förderlinie "Weiterentwicklung der IT-Infrastruktur(Software)" unterstützt. Darüber hinaus sind die seit vielen Monaten kon-

tinuierlich erbrachten Beiträge und Unterstützungen der Key User, Prozessverantwortlichen und aller Teilnehmern an den mehr als 40 Workshops hervorzuheben ohne deren Bereitschaft das Projekt nicht möglich wäre. Dank gilt auch den Dekanaten und der Universitätsleitung, welche die Rahmenbedingungen geschaffen haben, dass diese Mitarbeit möglich ist.

Literaturverzeichnis

[De09] Degenhardt, Lars; Gilch, Harald; Stender, Birga; Wannemacher, Klaus: Campus-Management-Systeme erfolgreich einführen. In (Hansen, Hans Robert; Karagiannis, Dimitris; Fill, Hans-Georg, Hrsg.): Wirtschaftsinformatik (2). Jgg. 247 in books@ocg.at. Österreichische Computer Gesellschaft, S. 463–472, 2009.

[DT03] Dougiamas, M.; Taylor, P. C: Moodle: Using learning communities to create an open source course management system. In: Proceedings of the EDMEDIA 2003 Conference, Honolulu, Hawaii. 2003.

[Ho13] Hortensius, Dick: Integrated management systems. Bericht, ISO Focus+, 2013.

[it] itelligence it.education. `http://itelligencegroup.com/de/industry-solutions/education/`. Accessed: 2015-04-18.

[Ja09] Janneck, Monique; Adelberger, Cyrill; Fiammingo, Sabine; Luka, Ruth: Von Eisbergen und Supertankern: Topologie eines Campus-Management-Einführungsprozesses. In (Hansen, Hans Robert; Karagiannis, Dimitris; Fill, Hans-Georg, Hrsg.): Wirtschaftsinformatik (2). Jgg. 247 in books@ocg.at. Österreichische Computer Gesellschaft, S. 453–462, 2009.

[Mo] Moodle Portal. `http://www.moodle.de`. Accessed: 2015-04-18.

[Ra09] Radenbach, Wolfgang: Integriertes Campus-Management durch Verknüpfung spezialisierter Standardsoftware. In (Hansen, Hans Robert; Karagiannis, Dimitris; Fill, Hans-Georg, Hrsg.): Wirtschaftsinformatik (2). Jgg. 247 in books@ocg.at. Österreichische Computer Gesellschaft, S. 503–512, 2009.

[SA] SAP SLcM. `http://www.sap.com/germany/solution/industry/higher-education-research.html`. Accessed: 2015-04-18.

[SAA12] Schreiter, Jan; Alt, Rainer; Auth, Gunnar: Business Engineering bei der Einführung von Campus-Management-Systemen - Herausforderungen und Potenziale. In (Goltz, Ursula; Magnor, Marcus A.; Appelrath, Hans-Jürgen; Matthies, Herbert K.; Balke, Wolf-Tilo; Wolf, Lars C., Hrsg.): GI-Jahrestagung. Jgg. 208 in LNI. GI, S. 642–656, 2012.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Strategisches zu Campusmanagementsoftware

André Kleinschmidt[1]

Abstract: Hochschulen sind stark heterogen organisierte Institutionen. Ihre Verwaltungsprozesse sind so vielgestaltig, wie deren Unterstützung durch IT-Systeme. Hersteller „voll integrierter" Campusmanagementsoftware (CMS) versprechen den Student Life Cycle umfänglich digital abzubilden. CMS sollen Einsparungen in Verwaltungen und Servicesteigerungen für Studierende herbeiführen.
Konferenzbeiträge um die Digitalisierung der Bildung lassen ausführliche Erfolgsgeschichten vermissen. Lassen sich CMS *(nicht)* reibungslos einführen? Werden erhoffte Vorteile erreicht? Erweisen sich unpassende CMS für Hochschulen als Fehlinvestitionen? Dieser Beitrag behandelt das Spannungsfeld digitaler Verwaltung unter verschiedenen Gesichtspunkten und liefert Anstöße für künftige Strategien und Entwicklungen.

Keywords: Campusmanagement, open/closed Software, digitale Verwaltung.

1 Einleitung

Die Einführung von Campus-Management-Software (CMS) geht mit monatelangen bis jahrelangen Projektphasen einher, unabhängig der Datenquellen (anderes CMS / analoges Archivmaterial) aus denen sie befüllt werden. Einige Projekte werden abgebrochen. In diesen Fällen bleiben gewohnte Arbeitsweisen unberührt, der Einsatz vorhandener CMS wird fortgeführt und eventuell wird erneut die Suche nach einem CMS ausgeschrieben. Hochschulleitungen haben nur während der Auswahlphasen der behandelten Systeme Einflussmöglichkeiten, vor Abschluss bindender Verträge mit Herstellern. Der vorliegende Beitrag soll auch den Entscheidern in solchen Phasen als Ideenquelle dienen, geeignete Festlegungen zu treffen.

Abschnitt 2 widmet sich potentiell auftretenden Komplikationen bei der Einführung von CMS, die Entscheider vor Planung von Einführungsphasen gezielt berücksichtigen können. Für die Einführung von voll integrierten CMS an Hochschulen finden sich verschiedene Interessenlagen auf die in Abschnitt 3 näher eingegangen wird. Insbesondere wird eine Gegenüberstellung der (wirtschaftlichen) Interessen der Softwarehersteller vorgenommen und ein strategisches Konfliktfeld aufgezeigt. Als Anregung für eine Abschwächung oder Aufhebung des beschriebenen Konfliktfeldes führt Abschnitt 4 alternative Ansätze zur Realisierung von IT-Unterstützung in Hochschulverwaltungen auf.

Abschnitt 5 fasst den Beitrag zusammen und ist als Empfehlung für Entscheider in Auswahlprozessen von CMS zu verstehen, dass den ausgewählten Punkten eine angemessene Aufmerksamkeit zu widmen ist.

[1]André Kleinschmidt, HHL Graduate School of Management, Department Information Technology, Officer Campusmanagement; andre.kleinschmidt(at)hhl.de

2 Komplikationen der CMS-Einführung

Voll integrierte CMS sollen Hochschulverwaltungen dabei unterstützen anfallende Geschäftsvorfälle abzubilden. Sie enthalten Berührungspunkte zu Kernprozessen der Auftragserfüllung (Forschung und Lehre) von Hochschulen bzw. der Wertschöpfung (private Träger).

Je größer der Kreis betroffener Akteure, deren Arbeit von CMS abgebildet werden soll, umso größer auch die negativen Effekte von Komplikationen bei der Einführung dieser Systeme. Vereinzelt werden Einführungsprojekte abgebrochen, bei Fachtagungen wird über solche Abbrüche zwar in informellen Gesprächen eingegangen. Für diesen Beitrag konnten keine öffentlich dokumentierten Fälle gefunden werden, die sich zitieren lassen.

Eine mögliche Ursache: Misserfolge und eine öffentliche Diskussion derer Gründe finden wenig Akzeptanz im Bereich wissenschaftlicher Veröffentlichungen. Zudem werden Einführungs-Projekte bzw. CMS im Produktivbetrieb teils unter Akronymen oder Projektbezeichnungen geführt, die vom Produktnamen der Hersteller abweichen, was die Recherche erschwert.

In den folgenden Unterabschnitten wird eine Auswahl von Gründen zur Anregung weiterer Diskussionen aufgezeigt. Probleme aus verschiedenen Ineffizienzen an Hochschulen waren bereits Gegenstand von Analysen [SKB10]. Die folgenden Unterabschnitte heben ausgewählte Spannungsfelder hervor.

2.1 Hochschulkultur wider Programmlogik

Gelebte Kultur in Hochschulen mit ihren selbst gelebten Freiheiten (in der Verwaltung) und gewährten Freiheiten (Studierenden gegenüber) ist nicht in formalen mathematischen Regeln auszudrücken. Trotz Bologna-geschuldeter Angleichungen und Standardisierungszwängen durch Massen von Studienbewerbern die derzeit und in Zukunft bewältigt werden müssen [JAE+14], funktionieren Hochschulen nicht wie Maschinen, sondern gleichen eher lebendigen Organismen, die einem ständigen evolutionsgleichen Weiterentwicklungsprozess folgen.

Hochschulverwaltungen verändern ihre Arbeitsweise und ihre Studienstrukturen je nach Notwendigkeit durch Änderungen von Rahmenbedingungen. Der Veränderungswille ist allerdings klassisch eher schwach ausgeprägt [SAA12]. Dem gegenüber stehen meist monolithische Softwaresysteme, die nicht auf häufige Veränderungen ausgelegt sind.

Es gibt Systeme, die veränderte Arbeitsabläufe berücksichtigen können; der größte Teil der Hersteller erwartet oder erzwingt gar eine Anpassung der Arbeitsweise der Verwaltungen an die realisierten Geschäftsvorfälle, so wie sie bereitgestellt werden. Entsprechend entwickeln sich Abwehrhaltungen seitens der Verwaltungen, die ihre gewohnten Arbeitsabläufe nicht aufgeben wollen oder können.

2.2 Verwaltungssprache ungleich Softwaresprache

Hochschulen und Hersteller bedienen sich eines unterschiedlichen Vokabulars. Nicht selten kann einem Terminus, der im Hochschulumfeld verwendet wird nicht bedeutungsgleich im CMS eines Herstellers wiedergefunden werden.

Studienordnungen (StudO) beispielsweise, setzen sich strukturell aus spezifischen Begriffen zusammen, wie Kompetenzbausteine [SOM14] oder Schlüsselqualifikationen [EU06]. Ein CMS hingegen erlaubt die mathematisch formelle Abbildung (Modellierung) von Studiengängen nur aus Elementen, die der Hersteller vorsieht: etwa Kurs, Modul oder Nebenfach. Diese Elemente entsprechen häufig nicht dem inhaltlichen Umfang der Hochschulterminologie.

Das gegenseitige „Sprache des anderen lernen“ nimmt viel Zeit in Anspruch, was mit der Zahlung von Beraterhonoraren einhergehen kann, die kalkulierte Kosten gemäß Ausschreibungen sprengen können. Auch laufen Gremien, die über Änderungen an Studienprogrammen entscheiden Gefahr, die Auswirkungen auf die Modellierbarkeit der Veränderungen im CMS nicht abschätzen können.

2.3 Einzelfälle wider Standardisierung

In einer „freien“ (= nicht an Softwareprozesse gebundene) Verwaltung sind Einzelabsprachen und Ausnahmen, wie Fristverlängerungen oder Härtefallregelungen ohne weiteres möglich sind. CMS wenden, die ihnen einprogrammierten formalen Regelungen häufig ohne Rückfragemechanismen „hart“ an.

Im zwischenmenschlichen Kontakt etwa zu einem Studiendezernat lässt sich eine Exmatrikulation wegen nicht erfolgter Rückmeldung (beispielsweise durch verzögerte Überweisung eines Semesterbeitrags) häufig noch durch „Augen zudrücken“ klären, während ein Studierender von einem automatisiert laufendem CMS bereits die Exmatrikulationsbescheinigung zugesandt bekommt.

Solche Regelverschärfungen - ohne Diskurs über rechtliche Fragen - sorgen insbesondere in der Studierendenschaft zu Abwehrhaltungen gegenüber voll integrierten Zentralsystemen.

2.4 Inkompatibilität zu bestehenden CMS

Wird ein CMS durch ein neues ersetzt, so soll häufig das abgelöste System abgeschaltet werden um Infrastruktur und Ressourcen zurückzugewinnen und um den Aufwand von Folgesupport für das nicht mehr genutzte System einzusparen. Dies bedingt je nach rechtlichen Grundlagen die Übernahme von Altdaten in das neue System um beispielsweise Auskunftspflichten (Abschlüsse, Noten, Bescheinigungen) für eine festgelegt Zahl von Jahren nachkommen zu können.

Es gibt in Deutschland keine standardisierten Datenformate zum Austausch von etwa Notendaten oder Ähnlichem. Softwarehersteller bieten häufig, meist textbasierte, Importschnittstellen an. Exportschnittstellen zur Ablösung von Systemen existieren hingegen kaum, da CMS von ihren Herstellern nicht für den Austausch konzipiert werden. Es liegt nicht im deren Interesse, die Ablösung des eigenen CMS zu erleichtern. Auswege können hier Datenkonvertierungen aus Schnittstellen für Berichtskomponenten bieten.

2.5 Begrenztheit auf den studentischen Lebenszyklus

Viele Hersteller werben damit den gesamten studentischen Lebenszyklus abzubilden. Diese Zentrierung auf die Studierenden greift teilweise zu kurz, wenn ein CMS den Anspruch erhebt die gesamte Verwaltung abbilden zu wollen. Viele wichtige Begleitprozesse, die in Hochschulverwaltungen anfallen, sind wenig oder gar nicht am Studierenden orientiert.

Als Beispiele lassen sich nennen:

- Bereitstellung von IT-Infrastruktur (E-Mailkonten, PC-Poolzugänge, Drucksysteme u.s.w.)
- Forecast von Lehrdeputaten
- Bibliotheksbetrieb
- Gebäude-, Havarie- und Wartungsmanagement sowie Materialbeschaffung

Diese Bereiche werden etwa von der Anzahl der Studierenden oder deren Nutzungsverhalten beeinflusst, aber selten sind sie in CMS integriert. Weiterhin stellt sich die Frage, was in ein CMS gehört und was nicht [BGS10]. Die Konsequenz der unvollständigen Abbildung der Verwaltung durch CMS ist die parallele Existenz verschiedener weiterer Systeme, mit denen Teilbereiche abgedeckt werden. Deren Anbindung an CMS gestaltet sich ebenfalls wegen fehlender Schnittstellen schwierig.

Als Gedankenspiel: Eine Havarie in einem Experimentalhörsaal macht diesen für den Rest des Semesters unbenutzbar. Die Havarie wurde von einer Gebäudemanagementsoftware erfasst. Wie fließt nun diese Information in das CMS, um eine Umplanung des Stunden- und Raumplans für betroffene Veranstaltungen anzustoßen? Wie werden die betroffenen Studierenden der Veranstaltungen systematisch informiert?

3 Interessen der Beteiligten

Im vorangegangenen Abschnitt wurde ein Beispiel für eine spezifische Interessenlage von CMS Herstellern genannt (keine Unterstützung der Abschaffung des eigenen Produkts). In den folgenden Unterabschnitten werden weitere benannt und ebenso die Interessen der Hochschulen als Software-Abnehmer umrissen. Werden die unterschied-

lichen Interessenlagen gegenübergestellt, zeigt sich ein Erklärungsansatz für Unzufriedenheiten mit CMS.

3.1 Interessen der Hochschulen

Ausschreibungen zur Anschaffung von Softwarelösungen sind aufwändig. Insbesondere Hochschulen in staatlicher Trägerschaft müssen Gesetze (z. B. SächsVergabeG [SVG13]), Ordnungen und je finanzieller Größenordnungen auch Europäische Richtlinien einhalten [EU15]. Die Idee voll integrierter CMS, die bei einmaliger Anschaffung möglichst viele Bereiche der Verwaltung abdecken zu können, kommt ausschreibenden Hochschulen entgegen.

Der Aufwand verschiedene Anforderungen (Software zur Studierendenverwaltung, Stunden- und Raumplanung, Prüfungsverwaltung etc.) auszuschreiben ist erscheint geringer als eine einmalige Ausschreibung, die alle Bedarfe zusammenfasst. Einzelne inhaltlich umfangreiche Ausschreibungen mit hohem monetären Auftragsvolumen sind häufiger anzutreffen, als spezifischere "kleinere" Ausschreibungen.

Hochschulen sind (öffentliche Träger) bzw. ähneln (private Träger) klassisch Behörden:

- Sie erfüllen öffentliche Aufträge (u.a. Forschung und Lehre).
- Hochschulen und ihre Organisationseinheiten (Fakultäten, An-Institute etc.) sind in ihrer Verwaltung autonom.
- Die Lenkung geschieht mittels (gesetzlich geregelter) Gremien.

Der gestiegene Wettbewerbsdruck zwischen den Hochschulen führt zu einer Verstärkung einer unternehmerischen Ausrichtung. In diesem Zuge steigt der Bedarf zur Datenerhebung von Kennzahlen als Grundlage für ein Controlling der Hochschulen.

Kennzahlen etwa über Studierenden- und Prüfungsstatistiken, Lehrendenauslastung, Gebäudenutzung etc. von dezentralen Systemen zu beziehen ist aufwändig. Der Einsatz von zentralen Systemen, die alle relevanten Daten der Hochschule vorhalten (single source of data) wirkt hier attraktiver als dezentrale Systeme, die gegebenenfalls gar nicht miteinander verbunden sind, etwa bei Fakultäten in Eigenverwaltung.

Für die meisten Hochschulen in Deutschland hat sich die finanzielle Situation durch Einsparungen in der Bildungspolitik verschlechtert. Einsparungen führen zu Stellenstreichungen, gleichzeitig soll jedoch nicht nur der Weiterbetrieb gewährleistet werden, sondern durch den Anstieg in der Bewerberlage zudem eine höhere Anzahl von Verwaltungsvorfällen (mit weniger Personal) bestritten werden. Dies führt zwangsläufig zur Bestrebung den Aufwand pro Vorfall zu verringern, etwa durch Standardisierung. Durch die Einführung von CMS werden Standardisierung und Arbeitserleichterung erhofft.

3.2 Interessen der Hersteller

Softwarehersteller sind in der Regel privatwirtschaftliche Unternehmen. Es gibt aber Ausnahmen, wie etwa die eingetragene Genossenschaft Hochschulinformationssysteme

(HIS) die als Selbstversorger ihrer Mitglieder fungiert [HIS14]. Unabhängig der Rechtsform müssen die Geschäftsmodelle der Hersteller auf dem Erwerb von Umsätzen basieren um insbesondere die Entwicklungskosten der vertriebenen CMS zu decken.

Der Absatzmarkt für CMS, gerade wenn sie sehr spezifisch an Deutsche Besonderheiten (z. B. dialogorientiertes Verfahren) angepasst sind, ist überschaubar. 425 Deutsche Hochschulen listet Destatis [DH15]. Nur etwa ein Duzend CMS-Anbieter bedienen den deutschen Markt. CMS stellen Nischenprodukte dar, die in einem bilateralen Oligopol gehandelt werden.

Weiterhin werden CMS nicht regelmäßig neu eingekauft. Die Wertschöpfung durch Verkauf ist unzureichend für die Aufrechterhaltung der Geschäftstätigkeit. Dauerhafte Einnahmen werden über Beratungs-, Support- und Einwicklungsleistungen erzielt, Geschäftsmodelle sind oft projektorientiert (vgl. [Dee10]).

3.3 Widersprüchliche Interessen

Aus den dargestellten Interessenlagen lassen sich Widersprüche ableiten, die eine qualitative Weiterentwicklung von CMS-Produkten bremsen und Erklärungsansätze für Probleme mit CMS-Einführungsprojekten sowie deren Produktivbetrieb bieten.

Die CMS-Hersteller müssen sich meist in wenigen offenen Ausschreibungen gegenüber Konkurrenten durchsetzen. Ein Entscheidungspunkt im Ausschreibungsprozess ist die Abdeckung der funktionalen Anforderungen (Pflichtenheft) durch Features des CMS. Dies führt gegebenenfalls zu einer Strategie der Produktentwicklung, die auf Quantität ausgerichtet ist, was zu Lasten der Qualität einzelner Features geht. Informeller ausgedrückt: Hersteller, die sich in Ausschreibungen durchsetzen wollen entwickeln eine hohe Anzahl von Features, aber ggf. keine „guten“ oder „gut durchdachten“.

Weiterhin fließen in die Produktentwicklungen die Anforderungen verschiedener Hochschulen ein, die – wie bekannt – heterogen arbeiten. Ein Feature das sehr spezifisch auf die Anforderungen einer Auftrag gebenden Hochschule (Kauf von Entwicklungsleistung) eingeht, führt zu Änderungen an den im CMS realisierten Geschäftsvorfälle, der grafischen Oberfläche, den Konfigurationsmöglichkeiten. Der Komplexitätsgrad der Software steigt zu Lasten ihrer Bedienbarkeit.

Ein Effekt, welcher den Interessen der Softwarehersteller zuspielt. Umsätze durch Verkauf von Schulungsleistungen sind wichtiger Bestandteil der Wertschöpfung der Unternehmen. Der Schulungsbedarf für Anwender ist für schwer verständliche und schwer bedienbare Software höher.

Eine Erhöhung der Bedienbarkeit, die Schaffung von Komfortfunktionen oder allgemeiner die Konzentration auf nicht-funktionale Anforderungen stellen für typische Geschäftsmodelle von Softwareherstellern keinen reizvollen Mehrwert dar.

Weiterhin ist zu betonen, dass die Produktentwicklung der Hersteller, die sich am Wertschöpfungsprozess orientiert und die Verwaltungsentwicklung der Hochschulen, die sich anhand wechselnder Rahmenbedingungen verändern, nicht miteinander rückgekoppelt sind. Sie verlaufen unabhängig voneinander.

Die Abwägung, ob ein angefragtes Feature seinen Weg in die Produktentwicklung beim Hersteller findet, richtet sich an der wirtschaftlichen Verwertbarkeit aus und weniger an der Dringlichkeit des Bedarfs auf Seiten der Hochschulen als Abnehmer.

Durch die Marktsituation des Oligopols und dem schwierigen und teuren Wechsel zwischen verschiedenen Anbietern, befinden sich Hochschulen, die sich einmal für einen Anbieter entschieden haben in einer gewissen Abhängigkeit von diesem.

Hochschulen haben wenig Druckmittel in der Hand um die Produktentwicklung der CMS in ihrem Sinne mitzubestimmen. Die Rechtsform der HIS eG als Versorger ihrer Mitglieder ist eine Seltenheit.

4 Veränderungsansätze

Die vorangegangenen Abschnitte haben das Spannungsfeld der Einführung und Nutzung von CMS in Hochschulen aus verschiedenen Perspektiven beleuchtet. Dieser Abschnitt soll Diskussionsgrundlage für eine gezielte strategische Veränderung des Spannungsfeldes dienen.

Die Geschäftsmodelle der Softwarehersteller werden sich nicht schlagartig ändern, selbst wenn Hersteller eine Notwendigkeit erkennen, diese mindestens zu überdenken. Ebenso können sich die Hochschulen nicht den vorliegenden Rahmenbedingungen (Auswahl CMS am Markt, Ablauf und Regelungen von Ausschreibungen etc.) entziehen.

Die vorgestellten Ansätze sind nicht neu, dennoch erscheint deren bewusste Berücksichtigung und ein fortgesetzter Diskurs wertvoll.

4.1 Monolithen zu Servicesystemen transformieren

Monolithische Softwaresysteme sind aufgrund ihrer hohen Komplexität schwierig zu warten und weiterzuentwickeln. Das Einfließen von Anforderungen verschiedener Abnehmer in ein standardisiertes Produkt, das an alle Kunden des Herstellers ausgeliefert wird, zieht verschiedene Implikationen nach sich:

- Steigerung von Konfigurationsaufwand und widersprüchliche Konfigurationsmöglichkeiten
- Senkung der Usability und Verständlichkeit
- Steigerung von Beratungs- und Schulungsaufwand

Als Alternativen zu monolithischen Anwendungen wird Konzepten wie Enterprise Application Integration (EAI), Serviceorientierte Architekturen (SOA) zunehmend Aufmerksamkeit gewidmet (Vgl. u.a. [Hof13], [AAG+10]).

Der noch junge Begriff des *Microservice* [Mic14] als *Untermenge des SOA-Begriffs* (Martin Fowler[2]) ist zu neu um zu bewerten, ob es sich um einen Hypebegriff handelt, oder dessen Konzepte sich erfolgreich etablieren.

Ein konzeptionelles Merkmal von Komponenten innerhalb des Microservice-Konzepts: Eine Komponente soll unabhängig ersetzbar und aktualisierbar sein. Dieser Punkt ist aus beiden Perspektiven (Hersteller, Abnehmer) interessant:

- Schneller Austausch im Fehlerfall.
- Möglichkeit zum Tausch einer Einzelkomponente bei Weiterentwicklung, außerhalb der (längeren) Releasezeiträume von komplexen Software-Monolithen.
- Die Möglichkeit zur Anfertigung spezieller Releases von Komponenten für spezifische Anforderungen, ohne Auslieferung an alle Kunden bei der Aktualisierung eines Software-Monolithen.

Im Zusammenhang mit Services existieren verschiedene Projekte, welche Daten aus verschiedenen Quellen - darunter die eingesetzten CMS – aggregieren. Auch über Frameworks in unterschiedlichen Entwicklungsstadien wird in diesem Zusammenhang publiziert [KGZ+14]. Die Verfolgung dieser Ansätze scheint lohnenswert. Besonders die häufig nicht zufriedenstellende Usability von CMS, kann durch ein Zwischenschalten (zwischen CMS und Endanwender) von Portalen und Apps verbessert werden kann.

Aktuelle Ansätze fokussieren häufig Dienste für Studierende und Lehrende, wie auch an die Showcase-Liste von StApps zeigt [PSA14]. Projekte wie *StApps* zielen nicht darauf ab, monolithische Softwaresysteme abzulösen, sondern darauf hin sie über Middlewares nutzbarer zu machen und aufzuwerten. Die Grundprobleme monolithischer Softwaresysteme lösen sie nicht.

4.2 Standards schaffen

Um bei der Entwicklung von CMS von monolithischen Systemen stärker auf Services zu orientieren, müssen die Schnittstellen zwischen Servicekomponenten harmonisiert werden. Ansonsten bleiben Entwicklungen wie Portale, Apps und Frameworks vereinzelte Silolösungen, die nur im Kontext weniger Hochschulen verwendet werden können, aber nicht von der Gemeinschaft aller Hochschulen.

Die Erkenntnis, dass Hochschulen heterogen organisiert sind und arbeiten steht dem entgegen. Informell ausgedrückt: *Die Option ist der Feind des Standards*. Um Softwarekomponenten, insbesondere jene aus der Hand verschiedener Hersteller, derart zu gestalten, dass sie miteinander interoperieren können, müssen Standards geschaffen werden.

Herausforderungen liegen dabei u.a. in folgenden Punkten:

---

2 Keynote zur GOTO Berlin 2014: https://youtu.be/wgdBVIX9ifA (formal nicht zitierfähig)

- Identifikation genügend typischer Gemeinsamkeiten als Grundlage
- Erweiterbarkeit für spezielle Anforderungen
- Schaffung von Anreizen zur Implementierung standardisierter Schnittstellen
- Findung einer geeigneten Institution zur Verabschiedung und Weiterentwicklung der Standards.

4.3 Eigenentwicklungen vorantreiben und offen teilen

Ein Konsens über gemeinsame Standards ermöglicht Hochschulen über Einzelprojekte, oder auch in Lehre integrierte Studierendenprojekte Eigenentwicklungen vorzunehmen, die ihre speziellen Anforderungen umsetzen.

Das Konzept von Systemen von Diensten, die auf Basis gemeinsamer Standards interagieren können deckt sich mit dem Community-Gedanken der OpenSource-Gemeinde und bricht darüber hinaus die Marktsituation des Oligopols teilweise auf.

Ein Möglicher Anreiz für die Bereitstellung von Referenzimplementierungen ist ein damit verbundener Prestigegewinn, wenn eine Komponente z. B. besonders nutzerfreundlich in ihrer Handhabe ist, oder ein Problem löst, das als Schwierig gilt (4-Ebenen Überschneidungsfreiheit von Stundenplänen, priorisierte Zuteilung von Studenten zu Projektgruppen).

4.4 CMS-Markt gezielt liberalisieren

Die Schaffung einer Auswahl von Komponenten, die Teilprozesse der Hochschulen besonders geeignet abbilden, kann langfristig zu einer Abschwächung des bilateralen Oligopols führen. Wenn Komponenten austauschbar sind, steigert das die Wettbewerbssituation. Ebenso können kleinere und mittlere Unternehmen Komponenten entwickeln, ohne die Notwendigkeit ein vollständiges CMS im eigenen Kostenrisiko bis zur Marktreife zu entwickeln.

Für etablierte Hersteller voll integrierter CMS, kann die strategische Orientierung hin zu Systemen von losen Komponenten, zu Beginn wenig attraktiv erscheinen: Die Öffnung für Standards erleichtert nicht nur die Platzierung des eigenen Produkts in einer standardisierten Umgebung sondern auch dessen Ablösung.

Insbesondere Hersteller, die sich durch monolithische Systeme mit breitem Verwaltungseinfluss die Abhängigkeit der Abnehmer sichern, ist zu erwarten, dass sie sich der Etablierung von Standards entgegenstellen. Hochschulen können eine Etablierung befördern, indem sie die Kompatibilität zu Standards bereits als Teil der Ausschreibung einfordern. Wenn ein bestimmtes CMS einmal eingeführt ist, kann es hierfür auf Grund der hohen Kosten der Einführungsprojekte zu spät sein.

5 Zusammenfassung und Ausblick

Der vorgelegte Beitrag, stellt bewusst mehr Fragen als er Antworten liefern soll. Die Aufzählung der Komplikationen bei CMS-Einführungen kann nur als Auswahl angesehen werden. Es ist anzunehmen, dass Hochschulen bei entsprechenden Einführungsprojekten mit einer Vielzahl von weiteren Problemen konfrontiert sind, die aus verschiedenen Gründen nicht öffentlich dokumentiert werden:

- Furcht vor Prestigeverlust
- Schweigevereinbahrungen mit CMS-Herstellern
- mangelnde Bereitschaft zur Veröffentlichung von Negativergebnissen

Die Beleuchtung der widersprüchlichen Interessen von Herstellern und Abnehmern von CMS sind nicht neu. Einige Leser werden ihnen lediglich den Charakter von Allgemeinplätzen einräumen. Die Einreichung dieses Beitrags im Workshop zielt auf die Steigerung von Bewusstsein ab, das bei Entscheidern in Ausschreibungen für CMS unterschiedlich ausgeprägt sein kann.

Gerade in Ausschreibungs-, Auswahl- und Entscheidungsprozessen ist der Widerspruch in den Interessenlagen besonders zu berücksichtigen. Nur in dieser Phase kann klar geregelt und eingefordert werden welche Rechte und Pflichten die zukünftigen Vertragspartner eingehen. Folgende Punkte werden den Entscheidern bei Auswahlprozessen für den Vertragsabschluss nahe gelegt:

- Regelungen über Folgeleistungen in Gewährleistung, Wartung und Support
- Realisierungszeiträume für Nachbesserungen unzureichender und Nachlieferung nicht vorhandener Features
- Klärung eines dauerhaften Produktverbesserungsabkommen
- Servicevereinbarungen über maximale Antwortzeiten bei Support-anfragen
- Kompensationsvereinbarungen bei Nichtabstellung von Fehlern

Nach Einführung eines CMS, wenn sich die Hochschulen in Abhängigkeit zu bestimmten Herstellern begeben haben, sind die Möglichkeiten zur Nachverhandlung begrenzt.

Die angegebenen Veränderungsansätze erheben keinen Anspruch auf Vollständigkeit. Sie sollen die weitere Diskussion und die Suche nach zufriedenstellenden Lösungen befruchten. Da sich die Rahmenbedingungen für Hochschulen auch in Zukunft weiter ändern werden, wird das Thema Campusmanagementsofware auf absehbare Zeit wenig an Relevanz verlieren.

Die derzeitige Situation für Hochschulen mit Blick auf verfügbare Systeme erscheint nicht zufriedenstellend. Die Verfolgung alternativer Ansätze und Strategien sollte mindestens aus Sicht der Hochschulen als lohnenswert erscheinen.

Literaturverzeichnis

[AAG+10] Addiks, J. S; Appelrath, H.-J.; Gringel, P.; Postina M.; Steffens, U.; Zillmann, C.: A Component-Oriented Tool-Approach to Enterprise Architecture Management. In Schumann, M.; Kolbe, L. M.; Breitner, M. H.; Frerichs, A. (Hrsg.): Multikonferenz Wirtschaftsinformatik 2010. Universitätsverlag Göttingen, , 2010.

[BGS10] Bick, M.; Grechening, T.; Spitta, T.: Campus-Management-Systeme. In Petisch, W.; Krams, B. Hrsg.(Hrsg.): Vom Projekt zum Produkt. Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik (WI-MAW). GI, Bonn, 61-78, 2010.

[Dee10] Deelmann, T.: Projekt- und produktorientierte IT-Unternehmen – Einige geschäftsmodellgestützte Überlegungen. In Schumann, M. et al.(Hrsg.): Multikonferenz Wirtschaftsinformatik. Universitätsverlag Göttingen, 87-92, 2010.

[DH15] Staat & Gesellschaft - Hochschulen - Statistisches Bundesamt (Destatis), https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Hochschulen/Hochschulen.html, Stand: 23.04.2015.

[EU06] EMPFEHLUNG DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 18. Dezember 2006 zu Schlüsselkompetenzen für lebensbegleitendes Lernen, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:394:0010:0018:DE:PDF, Stand: 20.04.2015.

[EU15] Public Contracts - Rules & procedures, http://europa.eu/youreurope/business/public-tenders/rules-procedures/index_en.htm, Stand: 23.04.2015.

[HIS14] HIS eG - Kurzportrait, https://www.his.de/ueber-his/kurzportrait.html, Stand: 23.04.2015.

[Hof13] Hofmann, J., Interoperabilität bei IT-Systemen imhochschulübergreifenden Kontext, [Dissertation] Universität Bremen, 2013

[JAE+14] Johnson, L.; Adams Becker, S.; Estrada, V.; Freeman, A., NMC Horizon Report: 2014 Higher Education Edition, Austin, Texas: The NewMedia Consortium, 2014

[KGZ+14] Kiy, A.;Grünewald, F.;Zoerner, D.; Lucke, U.: Ein Hochschul-App-Framework: Hybrid und modular. In Trahasch, S. et al.(Hrsg.): Die 12. e-Learning Fachtagung Informatik (DeLFI). Bonn : Köllen, 205-216, 2014.

[Mic14] Microservices, http://martinfowler.com/articles/microservices.html, Stand: 24.04.2015.

[PSA14] Generisches Framework Studi-App [PDF], https://www.ikmz.europa-uni.de/de/it-projekte/laufende/studi-app/ueber/2014-Projektskizze-Studi-App.pdf, Stand: 27.04.2015.

[SAA12] Schreiter, J.; Alt, R;Auth, G.: Business Engineering bei der Einführung von Campus-Management-Systemen - Herausforderungen und Potenziale. In Goltz, U.; Magnor, M.; Appelrath H.-J.; Matthies, H. K.; Balke, W-T.; Wolf, L. (Hrsg.): INFORMATIK 2012. Gesellschaft für Informatik, Bonn, 642-656, 2012.

[SKB10] Sprenger, J.;Klages, M.;Breitner, M.: Wirtschaftlichkeitsanalyse für die Auswahl, die Migration und den Betrieb eines Campus-Management-Systems. In Schumann, M. et al.(Hrsg.): Multikonferenz Wirtschaftsinformatik. Universitätsverlag Göttingen, 211-224, 2010.

[SOM14] Studienordnung Masterstudiengang Medieninformatik [PDF], http://www.htwk-leipzig.de/fileadmin/ordnungen/amtliche_bekanntmachungen/studien_und_pruefungs_ordnungen/Master/Medieninformatik/2014-06-20_online__StudO-MIM__manl_29102013_vorlage_RK_geschwaerzt.pdf, Stand: 20.04.2014.

[SVG13] Gesetz über die Vergabe öffentlicher Aufträge im Freistaat Sachsen (Sächsisches Vergabegesetzt - SächsVergabeG) vom 14. Februar 2013 [PDF], https://www.vergabe-sachsen.de/fileadmin/user_upload/pdf/SaechsVergabeG.pdf, Stand: 23.04.2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Systematische Prozessverbesserung mittels präferenzorientierter Ressourcenallokation am Beispiel einer Tutorienplatzvergabe

Andreas Drescher[1], Michael Meier[2], Andreas Oberweis[1] und Frederic Toussaint[2]

Abstract: Innovative IT-basierte Services im Hochschulsektor sind aufgrund der bereits ausgeprägten heterogenen und historisch gewachsenen IT-Landschaft mit zahlreichen individuellen IT-Lösungen kritisch zu betrachten. Eine Konsolidierung der heterogenen IT-basierten Services wird durch die Einführung von leistungsfähigen Campus-Management-Systemen sowie den Einsatz von eLearning-Systemen in der Lehre angestrebt. Somit sollte eine Verteilung von Studierenden auf Tutorien zum Standardrepertoire der Systeme gehören. Jedoch erfüllen die vorhandenen Lösungen nicht die aktuellen Ansprüche, wie beispielsweise eine einfache und intuitive Nutzung, eine optimale Gesamtzufriedenheit bezüglich der zugewiesen Termine für die Studierenden oder Schnittstellen zu bestehenden IT-Systemen. In diesem Zusammenhang hat die Fakultät für Wirtschaftswissenschaften des Karlsruher Instituts für Technologie das System YouSubscribe entwickelt, welches einen Algorithmus zur Verteilung von Studierenden auf Tutorien zur Verfügung stellt, mit dem eine maximale Gesamtzufriedenheit unter Berücksichtigung vieler Nebenbedingungen garantiert werden kann. Zu den Nebenbedingungen zählen beispielsweise Ressourcenvorgaben der Dozenten wie Minimal-, Maximal- oder Gleich-Auslastung von Räumen, Lerngruppenbildung oder Vermeidung der Zuteilung ungewünschter Tutorien. Durch vielfältige Schnittstellen mit datenführenden Systemen wird ein Medienbruch vermieden und die Datenintegrität gewährleistet sowie eine Integration in die Campus-Management-Systeme gewährleistet. Darüber hinaus wird für eine bestmögliche Medienunabhängigkeit ein Responsive Design eingesetzt.

Keywords: YouSubscribe, IT-basierte Services, Gesamtzufriedenheit, Studierende, Dozenten, Hochschule, Präferenzen, Tutorien, Prozessverbesserung, Ressourcenallokation, Optimierung.

1 Einleitung

Der Einsatz innovativer IT-basierter Services im Hochschulkontext, der beispielsweise in [KSL13, BRH14, Ka13, ZRL14, ZGL14, ZKB12] aufgezeigt wird, ist aufgrund der kontinuierlich steigenden Anzahl an Studierenden [St13] sowie durch die veränderten Arbeitsgewohnheiten der Studierenden [Br13] zu einer zentralen Aufgabe für Hochschulen geworden [He14]. An die geänderten organisatorischen und technischen Rahmenbedingungen muss die Prozess- sowie IT-Landschaft einer Hochschule angepasst werden. Im Rahmen dieser Anpassung ist es erstrebenswert, eine möglichst hohe Ergebniszufrieden-

[1] Karlsruher Institut für Technologie (KIT), Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB), 76128 Karlsruhe, Vorname.Nachname@kit.edu
[2] Karlsruher Institut für Technologie (KIT), Fakultät für Wirtschaftswissenschaften, Englerstraße 11, 76131 Karlsruhe, Vorname.Nachname@kit.edu

heit bei allen Prozessbeteiligten zu erreichen. Hilfreich ist eine systematische Prozessanalyse, um unter anderem Medienbrüche innerhalb der heterogenen und historisch gewachsenen IT-Landschaft zu identifizieren und durch innovative, integrierte IT-basierte Services zu unterbinden. In diesem Zusammenhang müssen zum einen die Aspekte des eLearnings [LS12] sowie zum anderen die Funktionen und Schnittstellen der Campus-Management-Systeme (CMS) [SAA12] berücksichtigt werden. Zusätzlich müssen die Anforderungen der Kunden, d.h. Studierende und Dozenten analysiert werden, um die Akzeptanz sowie die Usability der zu entwickelnden IT-Services sicher zu stellen [BB08].

In diesem Zusammenhang wurde an der Fakultät für Wirtschaftswissenschaften des Karlsruher Instituts für Technologie (KIT) der Bedarf nach einer präferenzorientierten Ressourcenallokation ermittelt. Insbesondere zu großen Lehrveranstaltungen im Grundstudium werden Tutorien angeboten, die zu unterschiedlichen Zeiten stattfinden und Studierende können aus der Anzahl an Tutorien auswählen (eine Lehrveranstaltung kann beispielsweise 900 Teilnehmer mit bis zu 50 verschiedenen Tutorienterminen aufweisen). Die Ressourcenallokation auf die Tutorien wurde vor der Einführung des IT-basierten Services YouSubscribe mit Hilfe von elektronischen Listen auf eLearning-Plattformen wie ILIAS (www.ilias.de) oder durch nicht-digitale Listen durchgeführt. Dementsprechend mussten sich Studierende in digitale oder nicht-digitalen Listen eintragen, welche nach dem First Come First Served – Prinzip funktionierten. Dies führte insbesondere bei nicht-digitalen Listen teilweise zu langen Schlangen und die Gesamtzufriedenheit der Studierenden konnte nicht sichergestellt werden.

In diesem Beitrag soll der IT-basierte Service YouSubscribe vorgestellt werden, welcher Dozenten und Studierende bei der präferenzorientierten Ressourcenallokation im Kontext der Tutorienplatzvergabe unterstützt. Mit Hilfe des Services soll eine optimale Gesamtzufriedenheit bei der Terminplatzvergabe gewährleistet werden. Dies bedeutet, dass möglichst alle Studierende Ihre Wunschtermine erhalten und keine mehrfachen Termine zur gleichen Zeit zugewiesen werden. Zusätzlich soll den Studierenden die Möglichkeit geboten werden, dass sie Lerngruppen bilden können und gemeinsam einem Tutorium zugeordnet werden. Der First-Come-First-Served – Ansatz wird durch eine „faire“ und transparente Terminbewertung ersetzt. Um das System möglichst gut in die vorhandene IT-Landschaft einzubetten, werden Schnittstellen zum Identity-Management, dem Vorlesungsverzeichnis und der eLearning-Plattform ILIAS verwendet.

2 Verwandte Arbeiten

Für die Allokation von Studierenden zu einer Vielzahl von Terminen im Hochschulkontext können unter anderem Module von eLearning-Plattformen, Campus-Management-Systeme oder spezialisierte Teilnehmer-Management-Systeme herangezogen werden. Im Folgenden soll auf eLearning-Plattformen sowie Campus-Management-Systeme eingegangen werden. Die spezialisierten Teilnehmer-Management-Systeme werden im nächsten Kapitel beschrieben.

Eine Marktstudie über eLearning-Plattformen hat weit über 100 proprietäre und nicht-proprietäre Produkte identifiziert [La15, BHM02]. eLearning-Plattformen unterstützen nicht nur Ihre Kernaufgabe, d.h. die Verwaltung und Organisation von Lernobjekten, sondern werden beispielsweise nach [Sc05] und [Gr09] durch eine Kurs- und Benutzerverwaltung, Analyse von Lernfortschritten, automatische Generierung und Auswertung von Tests sowie Werkzeugen zur Lehrplangestaltung angereichert. Sie bieten aber in der Regel keine präferenzorientierten Allokationsverfahren zur Sicherstellung einer optimalen Gesamtzufriedenheit für die Zuteilung von Studierenden zu einer Vielzahl von Terminen im Kontext der Tutorienplatzvergabe. Deutlich wird dies an der eLearning-Plattform ILIAS, welche unter anderem am KIT eingesetzt wird. Die Plattform ILIAS bietet eine Möglichkeit, Tutorienanmeldungen nach dem First Come First Served – Prinzip zu gestalten [IL15]. Dadurch findet zwar eine Gruppeneinteilung statt, jedoch kann keine optimale Gesamtzufriedenheit für die Terminwahl sichergestellt werden. Terminkollisionen bzw. Terminüberschneidungen muss der Studierende selbst kontrollieren und die Bildung von Lern- bzw. Zuteilungsgruppen wird nicht unterstützt. Aus diesem Grund erfüllt die eLearning-Plattform ILIAS nicht die notwendigen Anforderungen für eine präferenzorientierte Ressourcenallokation.

Campus-Management-Systeme sind integrierte Anwendungssysteme, welche die administrativen Prozesse einer Hochschule unterstützen und vereinheitlichen [SKB10, Br09, Be09, St07], aber auch Führungsinformationen [Kü98, Sp97] liefern. Nach der Studie [Er12] zählen die Produkte der HIS GmbH, CampusNet, SAP Student Lifecycle Management, CampusOnline und CAS Campus zu den bekanntesten. Der Funktionsumfang ist breit gefächert und unterstützt im Wesentlichen alle Hochschulprozesse, welche die Themenbereiche Forschung, Lernen, Lehre und den Student Lifecycle umfassen [Ca15]. Das KIT verwendet das Campus-Management-System CAS Campus, welches die Funktionsbereiche CRM und Interessenten, Bewerbung und Zulassung, Studierende und Gebühren, Studiengänge und Prüfungen sowie Veranstaltungen und Räume zur Verfügung stellt. Das Modul Veranstaltungen und Räume stellt beispielsweise ein Anmeldeverfahren für Tutorien auf Basis von Präferenzen zur Verfügung. Sofern mehrere Termine zur Verfügung stehen, können Studierende eine Rangfolge zwischen den Terminen bilden [CA15]. Hierbei ist es jedoch nicht möglich, mehrere Termine mit der gleichen Präferenz zu bewerten. Darüber hinaus wird keine Schnittstelle zu weiteren Veranstaltungen angeboten, so dass Terminkollisionen entstehen können. Dementsprechend besteht die Möglichkeit, dass ein Student zwei Tutorien zur gleichen Zeit besuchen müsste. Der Aspekt der Lerngruppenbildung sowie Schnittstellen zur eLearning-Plattform ILIAS werden ebenfalls nicht angeboten. Aus diesem Grund wäre es wünschenswert, ein leistungsfähigeres System in Bezug auf die Ressourcenallokation zur Verfügung zu stellen.

Spezialisierte Teilnehmer-Management-Systeme sollten eine bessere Lösung finden als eLearning-Systeme oder das vorgestellte Campus-Management-System. Deshalb werden im Folgenden einige vorhandene Systeme betrachtet und es wird die IT-basierte Plattform YouSubscribe vorgestellt, welche unter anderem die Aspekte der Terminkollision, Lerngruppenbildung, Raumressourcen und individuelle Gestaltung der Allokationszielfunktion inklusive entsprechender Nebenbedingungen betrachtet.

3 YouSubscribe

Vor der Beschreibung der Teilnehmer-Management-Systeme soll das eigentliche Problem genauer beschrieben werden.

3.1 Das Allokationsproblem und mögliche Lösungsalgorithmen

Am KIT werden Vorlesungen mit bis zu 900 Teilnehmern angeboten. Bei einigen Veranstaltungen können Studierende aus bis zu 50 begleitenden Tutorien dasjenige aussuchen, das am besten in ihren Terminplan passt. Die Auswahl per Einschreibelisten führte häufig zu Beschwerden bei den Studierenden, da die Wunschtermine bereits belegt waren. Mit Hilfe von YouSubscribe können die Studierenden Tutorien mit einem Zufriedenheitswert bewerten. Dadurch kann ein Maximierungsproblem $\sum_x \sigma_{ij} x_{ij}$ formuliert werden, bei dem x_{ij} die Bewertung von Student i für das Tutorium j ist. Durch die Indikatorvariable σ_{ij}, die bei jedem Studierendem genau einmal 1 ist, ansonsten 0, wird sichergestellt, dass jeder Studierende genau einen Platz erhält. Ohne weitere Nebenbedingungen ist das Problem leicht zu lösen. In der Realität besitzen aber Tutorien Maximalgrößen, die durch unterschiedliche Raumvorgaben für jedes Tutorium unterschiedlich sein können. Zusätzlich ist es sinnvoll, dass Tutorien Mindestteilnehmerzahlen besitzen (oder bei zu geringer Resonanz ganz ausfallen). Eine noch stärkere Nebenbedingung ist, dass alle Tutorien möglichst gleich stark belegt werden sollen, um eventuell Laborgeräte sinnvoll nutzen zu können. Soll zusätzlich noch berücksichtigt werden, dass sich Studierende zu Gruppen zusammenschließen dürfen, die gemeinsam ein Tutorium belegen möchten, wird das Problem NP-vollständig. Eine noch größere Herausforderung ergibt sich, wenn berücksichtigt werden soll, dass Studierende Tutorien in mehreren Veranstaltungen V belegen möchten und sich deren Termine nicht überschneiden sollen. Außerdem kann es sinnvoll sein, eine Gewichtungsfunktion ω vorzusehen, die absolut betrachtet identische Änderungen bei hoher Zufriedenheit anders bewertet als bei niedriger Zufriedenheit, so dass die Zielfunktion erweitert werden muss zu $\max_x \sum_V \sum_x \omega_{jV} \sigma_{ijV} x_{ijV}$ bei Beibehaltung der Nebenbedingungen.

Eine Aufgabe ist es, dieses mathematische Problem zu lösen. Die Layoutgestaltung für eine einfache und schnelle Dateneingabe für Studierende stellt ebenfalls eine Herausforderung dar. Zusätzlich müssen „unfaire und strategische Bewertungen“ möglichst unterbunden werden. Eine weitere Herausforderung ist es, wenn Dozenten ohne Erfahrungen in Optimierungsaufgaben eine schnelle, sichere und optimierende Zuteilung von Tutorien in einer Selbstbedienungsfunktion angeboten werden soll. Einige Hochschulen, wie beispielsweise die TU Berlin oder auch die RWTH Aachen verwenden als Teilnehmer-Management-Systeme die Werkzeuge MathPlan (www.mathplan.de) oder RedSeat (www.redseat.de). Auch am KIT gab es mehrere Tools, die für die Aufgabe eingesetzt wurden. Allen gemeinsam ist, dass sie in der Regel die Optimierung mit Heuristiken durchführen. Dadurch können sie über die Güte der Ergebnisse keine Aussage treffen. Bei keinem der genannten Werkzeuge ist den Autoren die Möglichkeit bekannt, eine Optimierung über mehrere Veranstaltungen hinweg durchzuführen. Sie bieten in der Regel auch

keine Schnittstellen zu den Campus-Management-Systemen, Identity-Management-Systemen oder eLearning-Plattformen. In den untersuchten Systemen können Dozenten unter anderem auch keine individuelle Zielfunktion für die Ressourcenallokation definieren. Aus diesem Grund wird nachfolgend die IT-basierte Plattform YouSubscribe vorgestellt.

3.2 Realisierung und Layoutgestaltung von YouSubscribe

In Abb. 1 wird der Prozess von der Erstellung einer Veranstaltung bis zur Publizierung der Ergebnisse illustriert, auf den nachfolgend detailliert eingegangen wird.

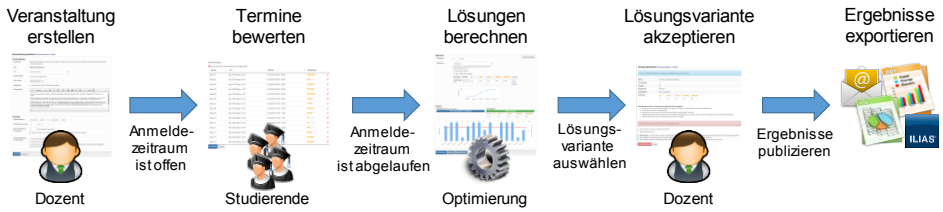


Abb. 1: Prozess des Teilnehmer-Management-Systems YouSubscribe

Im ersten Schritt wird der Dozent dazu aufgefordert, eine Veranstaltung anzulegen, wobei Veranstaltungen mit Hilfe des Vorlesungsverzeichnisses oder ohne Vorlesungsverzeichnisanbindung angelegt werden können. Durch die Schnittstellen zum Identity-Management sowie zum Modul Veranstaltungs- und Raummanagement des CAS Campus Systems werden alle Veranstaltungen aus dem Vorlesungsverzeichnis des KIT importiert, so dass ein vereinfachtes Anlegeverfahren ermöglicht wird, indem Rahmendaten sowie Rechte zu den Veranstaltungen bereits vorliegen. Ein Dozent erhält Rechte an einer Veranstaltung durch den Eintrag als Dozent oder Veranstaltungsbearbeiter im Vorlesungsverzeichnis. Zu den Rahmendaten zählen beispielsweise der Veranstaltungsname, die Veranstaltungsnummer, Kontaktdaten sowie die Einrichtung, wie beispielsweise Fakultät für Wirtschaftswissenschaften oder House of Competence. Neben den Rahmendaten einer Veranstaltung kann eine Beschreibung sowie ein Smart-Link erstellt werden. Darüber hinaus können die Rahmenbedingungen für das Anmeldeverfahren festgelegt werden, d.h. der Anmeldezeitraum, die minimale und maximale Teilnehmerzahl pro Tutorium auf globaler und lokaler Ebene, die maximale Größe einer Lerngruppe, Mindestanforderungen an den abzugebenden Bewertungsdurchschnitt für die Studierenden, relative Maximalanforderung für gleich zu bewertende Termine sowie die absolute Mindestanforderung an die abzugebenden Präferenzen, die Studierende aus Gründen der Einfachheit in einem 5-Star-Rating durchführen sollen (vgl. Abb. 2).

Nachdem die Rahmendaten einer Veranstaltung festgelegt wurden, können die einzelnen zu bewerteten Termine in das System eingetragen werden. Zusätzlich kann während und nach der Bewertung die Ansicht für eine aggregierte Übersicht über die Verteilung der abgegebenen Präferenzen herangezogen werden (vgl. Abb. 3). Durch diese Übersicht hat ein Dozent die Möglichkeit, auch während der Anmeldephase auf Unstimmigkeiten zu

reagieren und kann Termine bearbeiten, löschen, aber auch sofern der Bedarf besteht, weitere Termine hinzufügen.



Abb. 2: Rahmenbedingungen für eine Veranstaltung

Name	Ort	Termin	Min	Max	1★	2★	3★	4★	5★	Ø	Aktionen
Tutorium 1	20.13 HS001	Mo 08:00 - 09:30	-	60	593	34	21	10	19	1,3	
Tutorium 10	20.13 R006	Di 17:30 - 19:00	-	30	566	46	40	11	14	1,3	
Tutorium 11	20.13 HS001	Mi 08:00 - 09:30	-	60	413	73	80	50	61	1,9	
Tutorium 12	20.13 R006	Mi 08:00 - 09:30	-	30	419	71	76	50	61	1,9	
Tutorium 13	20.13 R109	Mi 14:00 - 15:30	-	40	285	27	38	43	284	3,0	

Abb. 3: Termine erstellen und aggregierte Bewertungen

Während des Anmeldezeitraums können Studierende die erstellen Termine mittels eines 5-Star-Ratings bewerten, wie in Abb. 4 illustriert. Studierende haben zusätzlich die Möglichkeit, Lerngruppen zu bilden, um sicherzustellen, dass alle Personen dieser Gruppe bei der Zuweisung denselben Termin erhalten. Jedoch ist dies nur möglich, wenn der Dozent die Option zugelassen hat (vgl. Abb. 2). Die Bewertungen können jeder Zeit durch die Studierenden während des Anmeldezeitraums verändert werden. Nach Ablauf des Anmeldezeitraums kann der Dozent die Studierenden mit Hilfe der einfachen oder angepassten Optimierung auf die Termine verteilen (vgl. Abb. 5). Um ein geeignetes Optierungsverfahren zu finden, wurden zu Beginn Untersuchungen mit Genetischen Algorithmen [Go04] und Tabu-Search-Verfahren [Gl89, Gl90] durchgeführt. Allerdings konnte weder die Laufzeit, noch die Ergebnisqualität überzeugen. Trotz Vorliegen eines NP-vollständigen Problems konnte der aus dem Operations Research bekannte Ansatz der linearen Optimierung erfolgreich evaluiert werden. Getestet wurde mit den Solvern lp_solve (http://lpsolve.sourceforge.net/5.5/), XPRESS (http://www.solver.com/) und Gurobi (http://www.gurobi.com/). Die Nebenbedingungen verhalten sich meist sehr gutmütig, so dass ein Dozent in der Regel innerhalb von 2-3 Sekunden ein nach vorgegebenen Kriterien optimales Ergebnis erhält. Lediglich bei einer sehr starken Nutzung von Lerngruppen und einer gewünschten Gleichauslastung aller Tutorien oder bei ungültigen Nebenbedingun-

gen wie zu wenigen Tutorienplätzen kann es zu unlösbaren Problemen kommen. Ungültige Nebenbedingungen findet der Algorithmus sofort. In diesem Fall muss der Dozent die Anzahl der Tutorienplätze erhöhen und/oder die Nebenbedingungen aufweichen. Sollte das Problem vom Solver innerhalb von 60 Sekunden nicht lösbar sein, wird die bis dahin beste Lösung als Ergebnis angeboten und der Dozent auf die Nicht-Optimalität hingewiesen. Dieser Fall ist allerdings bei den bisherigen realen Berechnungen nicht aufgetreten. Durch die angepasste Optimierung kann der Dozent eine individuelle Zielfunktion durch die Gewichtung der Bewertungen parametrisieren, sowie weitere optionale Nebenbedingungen, wie „keine leeren Termine“ zulassen oder „keine 1-Stern-Zuweisungen“ ergänzen. Bei der Terminkollisionsvermeidung über Veranstaltungen hinweg trat das Problem auf, dass es keinen Zeitpunkt gab, zu dem bereits Bewertungen zu allen Veranstaltungen abgegeben worden sind und nicht bereits für einzelne Veranstaltungen Lösungen benötigt wurden. Somit lag nie ein vollständiges Problem vor, das gelöst werden konnte. Um trotzdem Terminkollisionen zu vermeiden, wurden für alle Studierenden die zugewiesenen Termine einer Veranstaltung mit Terminen noch zu optimierender Veranstaltungen verglichen und im Fall einer Kollision deren Bewertungen heruntergesetzt. Da es vorkommen könnte, dass Studierende zwei Tutorien zum gleichen Termin besuchen wollen, werden die Studierenden über die Änderung per Mail informiert. Dann haben sie die Möglichkeit, die Werte wieder zu überschreiben.

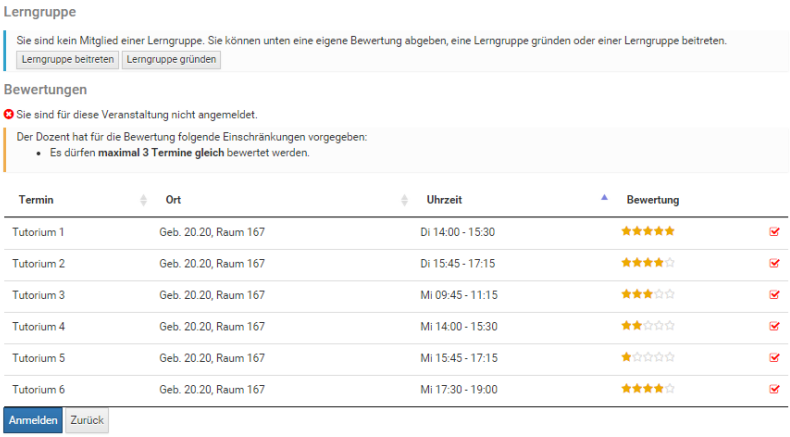


Abb. 4: Bewertung abgeben

Die Nebenbedingungen der Raumressourcen können vom Dozenten auf lokaler oder globaler Ebene jeder Zeit modifiziert werden. Ein geschicktes Vorgehen ist beispielweise, je Termin nur ein Tutorium bewerten zu lassen. Dadurch müssen die Studierende weniger Veranstaltungstermine bewerten. Vor der Auswertung werden die Ergebnisse für parallel liegende Termine einfach dupliziert (vgl. Abb. 6). Nach Beendigung der Berechnung, wird dem Dozenten die Auswertung grafisch aufbereitet. Zum einen wird eine relative Zufriedenheit unter Berücksichtigung der abgegebenen Bewertungen illustriert. Zum anderen

wird dem Dozenten ein kombiniertes Diagramm zur Verfügung gestellt, welches die Verteilung der Studierenden auf die Termine (Balkendiagramm), die relative aggregierte Zufriedenheit pro Termin (Punkte) sowie die durchschnittliche Anzahl der Studierenden pro Termin (Linie) darstellt (vgl. Abb. 7).

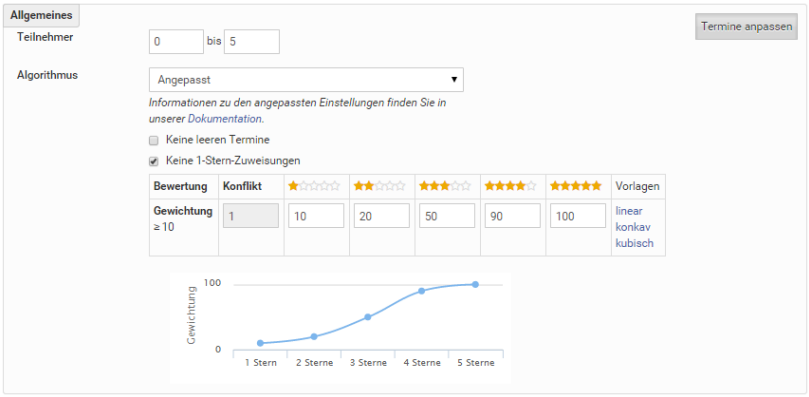


Abb. 5: Individuelle Zielfunktion und Nebenbedingungen



Abb. 6: Nebenbedingungen für Raumressourcen

Aufgrund der zahlreichen Parameter, die ein Dozent einstellen kann, bietet die Plattform die Möglichkeit, verschiedene Lösungen, namentlich Varianten, zu berechnen und abzuspeichern (vgl. Abb. 8), damit diese nachgelagert verglichen werden können. Im Rahmen der aggregierten Übersicht wird die jeweilige Gesamtzufriedenheit aufgezeigt. Für die Detailinformationen kann jede Variante individuell betrachtet werden. Sofern eine Variante den Anforderungen des Dozenten genügt, kann er diese akzeptieren (vgl. Abb. 8). Für die Publizierung der Ergebnisse wird dem Dozenten noch einmal eine Übersicht mit der Verteilung der Zuweisungen angezeigt (vgl. Abb. 9). Bei der Akzeptierung der Lösungsvariante werden Terminkonflikte identifiziert und wie oben beschrieben korrigiert. Für die Weiterverarbeitung oder Archivierung der Ressourcenallokation hat ein Dozent die Möglichkeit, die Ergebnisse als CSV- oder Excel-Datei zu exportierten. Weiterhin können die Ergebnisse direkt per Webservice in die eLearning-Plattform ILIAS überführt werden (vgl. Abb. 10). Hierbei wird im ILIAS-Kurs automatisch für jedes Tutorium eine Gruppe erstellt und die Teilnehmer jeweils dieser Gruppe zugewiesen. Sofern der Tutor dieser Gruppe hinzugefügt werden soll, muss dies aktuell noch manuell erfolgen.

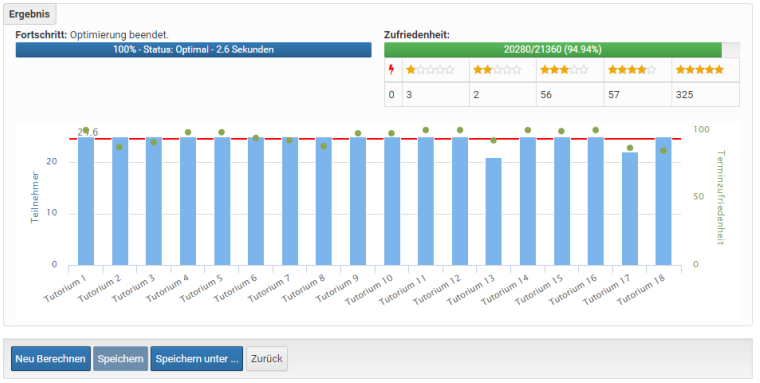


Abb. 7: Ergebnisübersicht

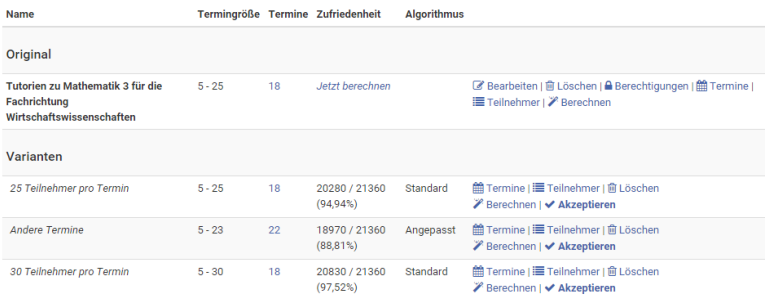


Abb. 8: Variantenauswahl

4 Zusammenfassung und Ausblick

Es wurde gezeigt, dass mit YouSubscribe ein Allokationsdienst angeboten werden kann, der den Dozenten in einer Selbstbedienungsfunktion die Möglichkeit bietet, komplexe Optimierungsverfahren in kurzer Zeit selbstständig durchführen zu lassen. YouSubscribe ist ein Dienst an der Fakultät für Wirtschaftswissenschaften am Karlsruher Institut für Technologie, um Dozenten bei der Verwaltung und Organisation einer optimalen Allokation von zahlreichen Studierenden zu einer Vielzahl von Terminen zu unterstützen. Bei ca. 6.000 Tutorienplatzvergaben je Semester liegt die Anzahl der Beschwerden unter 10, die Gesamtzufriedenheit je Veranstaltung liegt in der Regel zwischen 90% und 95%. Die optimale Verteilung basiert auf den Präferenzen der Studierenden und Dozenten. Studierende können beispielsweise Lerngruppen bilden, um sicherzustellen, dass sie in das gleiche Tutorium eingeteilt werden. Weiterhin haben Dozenten die Möglichkeit, individuelle Zielfunktionen sowie die dazugehörigen Nebenbedingungen zu konzipieren, um beispielsweise die zur Verfügung stehenden Raumressourcen optimal auszunutzen oder die gleichmäßige Auslastung von Tutorien sicherzustellen. YouSubscribe berücksichtigt auch den

Aspekt der Terminkollision, so dass gewährleistet wird, dass ein Student nicht mehrere Tutorien zum gleichen Zeitpunkt besuchen muss. Für eine bestmögliche Integration in die heterogene IT-Landschaft des KIT wurde die Plattform YouSubscribe durch entsprechende Schnittstellen mit dem Campus-Management-System CAS Campus, der eLearning-Plattform ILIAS sowie mit dem Identity-Management verknüpft. Die einzige Voraussetzung für die Nutzung von YouSubscribe ist eine Benutzerauthentifizierung, die durch eine Shibboleth-Anbindung, sowie ein Fallback mit Active Directory/LDAP realisiert wurde. Zusätzlich ist für externe Benutzer eine Authentifizierung über eine lokale Benutzerdatenbank realisiert. Alle weiteren Schnittstellen dienen der Steigerung der Benutzerfreundlichkeit und sind daher für den Betrieb nicht zwangsläufig erforderlich. Die Plattform wird kontinuierlich weiterentwickelt, um sie an die Bedürfnisse der Kunden anzupassen. Schwerpunkt ist hierbei, die Komplexität des Optimierungsvorgangs hinter einfach zu nutzenden Weboberflächen zu verbergen. Für eine bestmögliche Medienunabhängigkeit wird ein Responsive Design eingesetzt. Die Erfahrung hat gezeigt, dass trotz optimaler Ergebnisberechnung der Wunsch nach einer Tauschbörse für Tutorienplätze besteht. Auch möchten Dozenten gerne noch nach der Optimierungsphase Änderungen an der Terminzuweisung durchführen können.

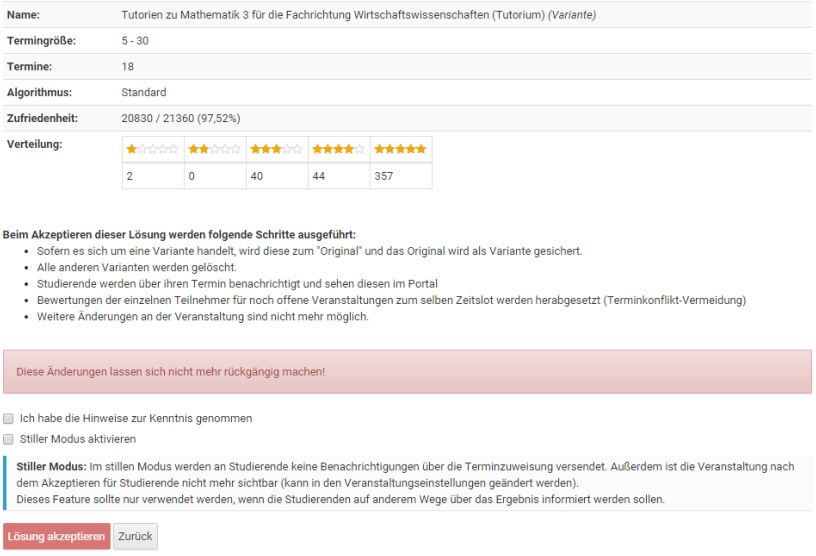


Abb. 9: Variante akzeptieren und Benachrichtigung

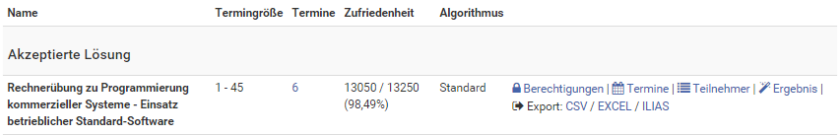


Abb. 10: Export

5 Literaturverzeichnis

[BB08] Bick, M.; Börgmann, K.: IT-Servicemanagement in deutschen Hochschulen. In HMD - Praxis Wirtschaftsinformatik, 2008; S. 105–113.

[Be09] Bensberg, F.: TCO-Analyse von Campus-Management-Systemen. Methodischer Bezugsrahmen und Softwareunterstützung. In (Hansen, H. R.; Karagiannis, D.; Fill, H.-G. Hrsg.): Business Services: Konzepte, Technologien, Anwendungen. 9. Internationale Tagung Wirtschaftsinformatik. Österreichische Computer Gesellschaft, Wien, 2009; S. 493–502.

[BHM02] Baumgartner, P.; Häfele, H.; Maier-Häfele, K.: E-Learning Praxishandbuch. Auswahl von Lernplattformen; Marktübersicht - Funktionen - Fachbegriffe. Studienverl., Innsbruck [u.a], 2002.

[Br09] Brune, H.; Jablonski, M.; Möhle, V.; Spitta, T.; Teßmer, M.: Ein Campus-Management-System als evolutionäre Entwicklung. Ein Erfahrungsbericht. In (Hansen, H. R.; Karagiannis, D.; Fill, H.-G. Hrsg.): Business Services: Konzepte, Technologien, Anwendungen. 9. Internationale Tagung Wirtschaftsinformatik. Österreichische Computer Gesellschaft, Wien, 2009; S. 483–492.

[Br13] Brugger, R.: A digital shift. In Switch Journal, 2013; S. 21.

[BRH14] Bittner, P.; Ritter, C.; Hildmann, T.: Get Your Study Plan. In (Plödereder, E. et al. Hrsg.): 44. Jahrestagung der Gesellschaft für Informatik. GI, Stuttgart, 2014; S. 1861–1872.

[Ca15] Carolla, M.: Ein Referenz-Datenmodell für Campus-Management-Systeme in deutschsprachigen Hochschulen. Springer, Wiesbaden, 2015.

[CA15] CAS Campus: Veranstaltungs- und Raummanagement. Anmeldeverfahren. http://www.cas-education.de/fuer-hochschulen/cas-campus/funktionen/veranstaltungen-raeume.html.

[Er12] Ernst & Young: Campus-Management – zwischen Hochschulautonomie und Bologna Reform, 2012.

[Gl89] Glover, F.: Tabu Search - Part I. In INFORMS Journal on Computing, 1989, 1; S. 190–206.

[Gl90] Glover, F.: Tabu Search - Part II. In INFORMS Journal on Computing, 1990, 2; S. 4–32.

[Go04] Goldberg, D. E.: Genetic algorithms in search, optimization, and machine learning. 26. Auflage. Addison-Wesley, Boston, 2004.

[Gr09] Gronau, N.: Wissen prozessorientiert managen. Methode und Werkzeuge für die Nutzung des Wettbewerbsfaktors Wissen in Unternehmen. Oldenbourg, München, 2009.

[He14] Heyde, M. von der: CIO-Modelle an deutschen Hochschulen. Netzwerk der entscheidenden Faktoren. In (Plödereder, E. et al. Hrsg.): 44. Jahrestagung der Gesellschaft für Informatik. GI, Stuttgart, 2014; S. 1815–1823.

[IL15] ILIAS open source e-Learning e.V.: ILIAS Dokumentation. http://www.ilias.de/docu/goto_docu_cat_581.html?lang=de, 2015.

[Ka13] Kaiser, S.; Kuhnt, E.; Lemcke, S.; Lucke, U.: Web-basierte Beschaffung. In (Horbach, M. Hrsg.): 43. Jahrestagung der Gesellschaft für Informatik. GI, 2013; S. 308–319.

[KSL13] Kiy, A.; Strelczuk, F.; Lucke, U.: Ein Plugin zur integrierten Literaturverwaltung in Moodle. In (Horbach, M. Hrsg.): 43. Jahrestagung der Gesellschaft für Informatik. GI, 2013; S. 334–345.

[Kü98] Küpper, H.-U.: Struktur, Aufgaben und Systeme des Hochschul-Controlling. In (Küpper, H.-U.; Sinz, E. J. Hrsg.): Gestaltungskonzepte für Hochschulen. Effizienz, Effektivität, Evolution. Schäffer-Poeschel, Stuttgart, 1998; S. 152–172.

[La15] Lang, M.: Learning Management Systeme (LMS) im Vergleich: Open Source-Lösungen oder proprietäre Produkte? 1. Auflage. disserta Verlag, Hamburg, 2015.

[LS12] Lucke, U.; Schroeder, U.: Forschungsherausforderung des E-Learning. In i-com, 2012, 11; S. 1–2.

[SAA12] Schreiter, J.; Alt, R.; Auth, G.: Business Engineering bei der Einführung von Campus-Management-Systemen. Herausforderungen und Potenziale. In (Goltz, U. et al. Hrsg.): 42. Jahrestagung der Gesellschaft für Informatik. GI, 2012; S. 642–656.

[Sc05] Schulmeister, R.: Lernplattformen für das virtuelle Lernen. Evaluation und Didaktik. 2. Auflage. Oldenbourg, München [u.a.], 2005.

[SKB10] Sprenger, J.; Klages, M.; Breitner, M. H.: Wirtschaftlichkeitsanalyse für die Auswahl, die Migration und den Betrieb eines Campus-Management-Systems. In Wirtschaftsinformatik, 2010, 52; S. 211–224.

[Sp97] Spitta, T.: Standardsoftware zur Verwaltung und Führung von Fakultäten. Bestandsaufnahme und Anforderungen. Diskussionspapier Nr. 354. Faculty of Economics and Business Administration, Bielefeld University, 1997.

[St07] Stender, B.; Jablonski, M.; Brune, C.-H.; Möhle, V.: Campus Management von der Hochschule aus gedacht. Werkstattbericht aus der Universität Bielefeld. In Wissenschaftsmanagement: Zeitschrift für Innovation, 2007, 2007; S. 19–26.

[St13] Statistisches Bundesamt: Studierende. https://www.destatis.de/DE/ZahlenFakten/Indikatoren/LangeReihen/Bildung/lrbil01.html?cms_gtp=152374_list%253D1&https=1, 01.04.2015.

[ZGL14] Zoerner, D.; Gößler, F.; Lucke, U.: Hochschul-Apps im Überblick. In (Plödereder, E. et al. Hrsg.): 44. Jahrestagung der Gesellschaft für Informatik. GI, Stuttgart, 2014; S. 1849–1860.

[ZKB12] Zakhariya, H.; Kosch, L.; Breitner, M. H.: Elektronische Drittmittelakte in der Hochschulverwaltung. Erkenntnisse aus Fallstudien. In (Goltz, U. et al. Hrsg.): 42. Jahrestagung der Gesellschaft für Informatik. GI, 2012; S. 613–626.

[ZRL14] Zender, R.; Rogozhkin, V.; Lucke, U.: Einsatz von MultiTouch-Tischen zur kooperativen Semesterorganisation. In (Plödereder, E. et al. Hrsg.): 44. Jahrestagung der Gesellschaft für Informatik. GI, Stuttgart, 2014; S. 1873–1884.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Modulempfehlungen über Vorgänger- und Nachfolgemodule

Patrick Bittner[1]

Abstract: Das moderne Studium stellt die Studierenden vor inhaltliche, organisatorische und technische Herausforderungen. Nicht jeder Studiengang hat einen klaren Ablauf und selbst wenn ein solcher vorhanden ist, ist er oft auf einen Studierenden ausgelegt, der kein Modul wiederholen muss. Außerdem können meist nicht alle Lebensentwürfe und –umstände berücksichtigt werden. In diesem Beitrag wird ein Konzept für ein Empfehlungssystem vorgestellt, welches aufgrund von Studienverläufen Abhängigkeiten zwischen den Modulen identifiziert. Hierbei werden sowohl die Voraussetzungen für die Module als auch mögliche Präferenzen der Studierenden bei einer Empfehlung berücksichtigt.

Keywords: Empfehlungssysteme, Studieplan-Optimierung, recommender system, Modulauswahl

1 Einleitung

An den Universitäten ist IT-Unterstützung heute nicht mehr wegzudenken. Den Studierenden werden unterschiedliche Dienste angeboten, wie beispielsweise Online-Bewerbungen oder E-Mail-Accounts.

Viele Studierende bringen bereits zum Studienanfang ihre eigene IT-Umgebung mit und verknüpfen die universitären Dienste mit ihrer eigenen Infrastruktur. Diese Entwicklung wird von den Universitäten gefördert, um den Studierenden ein größtmögliches Maß an Komfort zu bieten.

Trotzdem ist der Verwaltungsaufwand während eines Studiums groß. Unterschiedliche Studienordnungen schaffen Regeln, an die sich die Studierenden der unterschiedlichen Fachrichtungen zu halten haben. Module müssen gesichtet und einzelne Vorlesungen ausgewählt werden, um die nötigen Anforderungen letztlich erfüllen zu können. Dabei ist es nicht immer einfach, die genauen Anforderungen der Prüfungsordnung oder auch nur eines bestimmten Moduls zu erfassen. Selbst Studiengänge, die einen relativ klaren Studienverlaufsplan vorschlagen, können nicht weiterhelfen, wenn z.B. eine Prüfung nicht bestanden wurde oder nur ein Teil der Prüfungen angetreten werden konnte. Bisher existieren wenige Systeme, die es den Studierenden leichter machen, herauszufinden, welche Module von ihnen noch zu belegen sind, und welche Voraussetzungen dafür mitzubringen sind. So müssen sich die meisten Studierenden selbst informieren, was ihnen überhaupt noch zum Abschluss ihres Studiums fehlt.

[1] Technische Universität Berlin, tubIT – IT-Service-Center der TU-Berlin, Einsteinufer 17, 10587 Berlin, patrick.bittner@tu-berlin.de

Für die Entwicklung eines Systems, welches die Studierenden an dieser Stelle unterstützt, müssen die Rahmenbedingungen in Form von Studienordnungen und Modulhandbüchern bekannt sein. Um allerdings tatsächlich auf einzelne Studierende abgestimmte Empfehlungen geben zu können, muss dem System bekannt sein, welche Vorlieben der Studierende hat, und welche der möglichen Module diesen Vorlieben am ehesten entgegenkommen. Dabei können beispielsweise die Studienverläufe anderer Studierender herangezogen werden. Je ähnlicher sich die Studierenden sind, desto wahrscheinlicher erscheinen ähnliche Modulvorlieben.

In dieser Arbeit werden wir zunächst bisherige Unterstützungsansätze für Studierende vorstellen, sowie gängige Empfehlungsansätze und Probleme bei solchen benennen. Anschließend stellen wir das Konzept für ein Empfehlungssystem für Module vor, welches auf den bisherigen Belegungshistorien der Studierenden basiert, und über Belegungszusammenhänge Empfehlungen generiert. Abschließend geben wir einen Ausblick über mögliche, weitere Ausbaumöglichkeiten dieses Systems.

2 Verwandte Arbeiten

Es existieren bereits einige Systeme, die den Studierenden bei der Organisation ihres Studiums helfen sollen. Im Folgenden werden beispielhaft zwei derartige Systeme genannt, die sich bereits im Einsatz befinden.

[FH09] beschreibt die Einführung eines Campus Management Systems in der Universität Hamburg. Das System beinhaltet auch die Optimierung der Modulplanung, um zeitliche Überschneidungen zu minimieren. Auf diese Weise können Studierende Module, welche für das gleiche Semester geplant sind, auch im gleichen Semester belegen. Dabei liegt der Schwerpunkt aber auf den Zeit- und Raumbelegungen. Eine Information an die Studierenden, welche Module für sie empfehlenswert sein könnten, erfolgt auf diese Weise jedoch nicht.

In [Ri09] wird ein im System integriertes „Self Monitoring“ angesprochen. Dieses ermöglicht es den Studierenden, sich selbst einzuschätzen, indem sie ihre Noten im Vergleich zu anderen Studierenden einsehen können. Ein zusätzliches „Self Assessment“ dient zur besseren Einschätzung der eigenen Stärken und Schwächen und berücksichtigt dabei den bisherigen Studienverlauf. Die Daten werden nicht dazu verwendet, konkrete Modulempfehlungen zu generieren. Unser Ansatz hingegen hat genau dieses Ziel, verzichtet aber bisher auf die Bereitstellung eines Tools zur Selbsteinschätzung.

2.1 Verbreitete „Recommender“-Ansätze

Um Empfehlungen zu generieren gibt es diverse unterschiedliche Ansätze, die seit langer Zeit Thema der Forschung sind. In [GT05], [He04], [Bu02] und [SK09] sind einige davon beschrieben.

Die meisten Ansätze für Empfehlungssysteme lassen sich in eine der folgenden Kategorien einordnen:

- *Demographische Ansätze* versuchen, über die statischen Eigenschaften der User Ähnlichkeiten zu finden. Dabei wird angenommen, dass User mit ähnlichen Eigenschaften ähnliche Präferenzen haben, woraus Empfehlungen generiert werden können.

- Bei *Inhalts- (Content-) basierten Ansätzen* werden Ähnlichkeiten zwischen den Eigenschaften der Items gesucht. Die Annahme besteht hierbei, dass User Items der gleichen „Art" bevorzugen.

- *Wissens- (Knowledge-) basierte Ansätze* verwenden ebenfalls die Eigenschaften der Items, besitzen aber zusätzlich das Wissen darum, was der User konkret braucht. Dieses Wissen kann z.B. über konkrete Abfragen gewonnen werden.

- *Collaborative Ansätze* gehen davon aus, dass User, die bisher ähnliche Präferenzen hatten, also ähnliche Bewertungen abgegeben haben, auch weiterhin die gleichen Items bevorzugen.

- *Modellbasierte Ansätze* versuchen, über Trainingsdaten Muster im Wahlverhalten zu erkennen und auf diese Weise Empfehlungen zu generieren. Dabei werden verschiedene Techniken verwenden, wie beispielsweise Matrix-Faktorisierung oder Bayessche Netze.

Je nach Ansatz ist ein Empfehlungssystem mit unterschiedlichen Herausforderungen konfrontiert. Es gibt einige gängige Probleme im Zusammenhang mit Empfehlungssystemen:

- Das *New-User Problem* bezeichnet die Problematik beim Anlegen eines neuen Users. Da ein neuer Benutzer anfangs noch keine oder nur sehr wenige Bewertungen abgegeben hat, können noch keine Präferenzen identifiziert werden. Erst wenn die Vorlieben des Users besser bekannt sind, können für ihn passende Items empfohlen werden.

- Als *New-Item Problem* bezeichnet man die analoge Problematik beim Anlegen eines neuen Items. Hier können zwar Item-Eigenschaften beim Erstellen bekannt sein, aber auch hier gibt es anfangs keine oder nur wenige Bewertungen. Dadurch können diese neuen Items in den Ansätzen, die Präferenzen berücksichtigen, nicht als Empfehlungen identifiziert werden, bis genug Daten vorhanden sind.

- Das *Cold-Start Problem* ist hierbei noch allgemeiner. Es bezeichnet den Umstand, dass ein System erstellt wird, aber keinerlei Bewertungen, also Verknüpfungen zwischen Usern und Items, vorhanden sind. Neu erstellte

Systeme ohne vorhandene Bewertungen können keinerlei Präferenzen identifizieren. Da jegliche Präferenzen in diesem Fall unbekannt sind, ist eine Generierung von Empfehlungen durch viele Ansätze so nicht möglich, bis eine Datenbasis geschaffen wurde.

Eine gängige Praxis, um den genannten Problemen entgegenzuwirken, ist die Hybridisierung der Ansätze. Hier wird versucht, die Ansätze zu kombinieren. Dadurch kann beispielsweise der Effekt fehlender Daten minimiert werden, um auf diese Weise trotzdem eine möglichst adäquate Empfehlung abgeben zu können. In [Bu02] werden einige Hybridisierungsansätze untersucht und verglichen.

3 Empfehlungsansatz

An den Universitäten wird eine Vielzahl an unterschiedlichen Modulen angeboten, welche sich wieder in verschiedene Veranstaltungen aufteilen können. Allerdings ist nicht jedes Modul für jeden Studiengang sinnvoll oder auch nur verwendbar. Eine erste Vorauswahl der Module ist also bereits durch eine Berücksichtigung der Studienordnung möglich. Das heißt, dass für einen Studierenden nur Module vorgeschlagen werden können, wenn diese ihn entsprechend seiner Studienordnung überhaupt weiterbringen können.

Diese Einschränkung ist jedoch sehr allgemein und bringt dem Studierenden kaum einen Mehrwert, da er sich dessen im Allgemeinen bereits bewusst ist.

Um eine auf einen einzelnen Studierenden abgestimmte Empfehlung geben zu können, muss zuerst klar sein, was über den Studierenden und über die Module überhaupt bekannt ist.

Im Folgenden stellen wir einen eigenen Ansatz vor und betrachten die Stärken und Schwächen auch im Vergleich zu anderen Ansätzen.

3.1 Modulhistorien der Studierenden als Input für das Empfehlungssystem

Um sein Studium abzuschließen, muss ein Studierender Module wählen und belegen. Daher ist ein Empfehlungssystem, welches im universitären Umfeld Module vorschlägt, nicht mit einem System zu vergleichen, welches zum Beispiel Kaufempfehlungen ausspricht. Der Zwang, tatsächlich etwas wählen zu müssen, sowie der Umstand, dass sämtliche bereits gewählten Module bekannt sind, unterscheiden diese spezielle Domäne von vielen anderen.

Die bisherigen Modulbelegungen der Studierenden sind als Datenbestand vorhanden. Über diese Belegungsdaten lässt sich eine Art Historie für jeden einzelnen Studierenden erzeugen, also eine Aufstellung, in welchem Semester ein Studierender welches Modul belegt hat. Diese Historien können als Input verwendet werden, um Empfehlungen zu

generieren. Ein Studierender im ersten Semester würde hier mit einer leeren Historie beginnen.

3.2 Vorgängermodule

Unser Ansatz geht davon aus, dass die Reihenfolge, in der Module gewählt werden, von den Präferenzen der Studierenden und deren Wissensstand abhängig ist. Ein bestimmtes Modul wird gewählt, wenn es durch die Studienordnung vorgegeben wird oder der Studierende ein Interesse für das Modul aufweist. Der Zeitpunkt der Belegung ist davon abhängig, ob der Studierende bereits alle nötigen Vorkenntnisse hat. In Modulbeschreibungen sind oftmals bereits solche Anforderungen an die Studierenden hinterlegt, aber nicht immer lässt sich damit das gesamte Spektrum abdecken.

In diesem Ansatz analysieren wir die Studierendendaten, um herauszufinden, welche Module als Vorgänger für welche anderen Module angesehen werden können. Dabei wird für jedes Modul festgehalten, wie viele Studierende dieses Modul gewählt haben, und welche Module von wie vielen Studierenden gewählt wurden, bevor das untersuchte Modul gewählt wurde. Über den Anteil der Studierenden, die ein bestimmtes Modul vorher gewählt haben, lässt sich das Ausmaß der Verbindung zwischen den Modulen abschätzen. Ein Modul, welches von einem großen Teil der Studierenden vor einem anderen gewählt wurde, kann als potentielles Vorgängermodul angesehen werden. Dazu wird ein Minimalanteil festgesetzt. Übersteigt der Anteil der Studierenden, welche ein Modul vor dem untersuchten Modul besucht haben, diesen Minimalanteil, so gilt das Modul als Vorgängermodul für das untersuchte Modul.

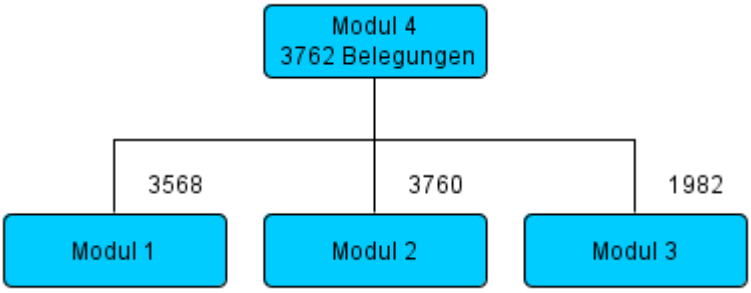


Abbildung 1: beispielhafte Darstellung des Moduls 4 mit seinen Vorgängermodulen

Wir gehen davon aus, dass man, wenn man sämtliche Vorgängermodule belegt hat, potentiell auch an dem eigentlichen Modul interessiert ist und auch die nötigen Kenntnisse erworben hat, um dieses Modul zu belegen.

Im Beispiel aus Abbildung 1 wird das Modul 4 von 3.762 Studierenden belegt. Bei der Betrachtung des Moduls wird festgestellt, dass laut den Historien der Studierenden, die

das Modul belegt haben, drei Module vor dem Modul 4 gewählt wurden. Diese drei Module sind potentielle Vorgängermodule. Während fast 100% (3760 von 3762) der Studierenden, die das Modul gewählt haben, vorher das Modul 2 gewählt haben, war der Anteil derer, die vorher das Modul 3 wählten, nur bei rund 55%. Das Modul 1 wurde von ca. 95% der Studierenden vor Belegung des Moduls 4 besucht. Je nach Parametrisierung, also Wahl des Minimalanteils, können die Module 1-3 nun als Vorgängermodul für das Modul 4 identifiziert werden.

3.3 Kombinationsvorgänger

Die vorgestellte Vorgängerbeziehung stellt einen direkten Zusammenhang zwischen zwei Modulen her. Diese zeigt an, dass ein Großteil der Studierenden vor dem untersuchten Modul das Vorgängermodul belegt hat. Es erscheint daher sinnvoll, Module nur zu empfehlen, wenn vorab sämtliche Vorgängermodule belegt wurden.

Allerdings können auch Module, die den in 3.2 angesprochenen Minimalanteil an Studierenden nicht erreichen (und somit keine Vorgängermodule sind) entscheidend dafür sein, ob ein Modul empfohlen werden sollte. Dies wird z.B. deutlich, wenn in mehreren Modulen ähnliche Grundlagen vermittelt werden. Wenn nun ein Modul auf diesen Grundlagen aufbaut, wird erwartet, dass mindestens eines der Grundlagenmodule gewählt wurde. Diese tauchen aber ggf. nicht als Vorgängermodul auf, da der Anteil an Studierenden, die das Grundlagenmodul vor dem untersuchten Modul belegt haben, ggf. recht klein ist.

An dieser Stelle kann eine weitere Analyse helfen, die feststellt, ob ein kombinierter Zusammenhang aus mehreren Modulen besteht. Darüber lassen sich Aussagen der Art „es ist sinnvoll, mindestens ‚X' dieser Module vor Antritt des Moduls belegt zu haben" treffen, die auch zur Einschränkung der Empfehlungen verwendet werden können.

3.4 Nachfolgemodule

Analog zur Analyse der Module auf Vorgänger ist auch eine Überprüfung auf Nachfolger denkbar. Diese wäre analog zur Vorgänger-Analyse aufzubauen.

Die Aussagen von Nachfolger- und Vorgängerbeziehungen sind hierbei nicht als redundant zu betrachten. Während eine Vorgängerbeziehung Auskunft darüber gibt, dass ein Großteil der Studierenden das Vorgängermodul bereits belegt hat, sagt sie nichts darüber aus, wie viele der Studierenden, welche das Vorgängermodul besucht haben, sich anschließend für das betrachtete Modul entschieden haben. Diese Lücke wird durch die Analyse auf Nachfolge geschlossen. Dabei ist zu erwarten, dass die Nachfolge-Analyse tendenziell die Präferenzen der Studierenden aufdeckt, während die Vorgänger-Analyse sich auch für die Aufdeckung von Voraussetzungen eignet.

Eine Kombination aus Vorgänger- und Nachfolgeanalyse erhält somit ein Potential,

fundierte Empfehlungen generieren zu können.

3.5 Vor- und Nachteile

Der von uns vorgestellte Empfehlungsansatz arbeitet mit dem Umstand, dass die Modulwahl als Sequenz aufgebaut ist, also eine Reihenfolge besitzt. Kein Modul kann mehrmals (erfolgreich) belegt werden und im Regelfall wird keine explizite Bewertung vorgenommen. Die angesprochene Reihenfolge wird von unserem Ansatz aufgegriffen, um daraus direkt Empfehlungen zu generieren. Auf diese Weise können fachliche und interessenmäßige Zusammenhänge erkannt und berücksichtigt werden. Außerdem bietet sich die Möglichkeit, die Module zu analysieren, wodurch die Lehrkräfte in die Lage versetzt werden, die von ihnen angebotenen Module noch besser im Gesamtkontext einschätzen zu können.

Der Ansatz baut auf den bisherigen Modulbelegungen auf. Sind diese nicht vorhanden, können keine Vorgänger oder Nachfolger-Beziehungen zwischen den Modulen identifiziert werden. Der Ansatz ist also anfällig für das in 2.1 beschriebene Cold-Start Problem.

Das New-User Problem betrifft diesen Ansatz nicht. Ein neuer Studierender ist ein User mit leerer Historie. Diesem User können direkt Module vorgeschlagen werden, welche keine Vorgänger besitzen oder der Nachfolger von keinem Modul sind. Dies sind im Wesentlichen die Module, welche im ersten Semester gewählt werden, und somit auch tendenziell für neue Studierende geeignet sind.

Bei Verwendung unseres Ansatzes in Reinform wird ein neu erstelltes Modul vom System nicht empfohlen werden. Dies begründet sich dadurch, dass das Modul in diesem Fall noch von niemandem gewählt wurde, und somit kein Vorgänger oder Nachfolger ist.

Wir gehen davon aus, dass in einer Hochschule die Studienverlaufspläne der einzelnen Studierenden vorliegen und als Daten verwendet werden können. Somit ist das Cold-Start Problem für den vorgestellten Ansatz nur von theoretischer Bedeutung: Unter Berücksichtigung des Datenschutzes sind die bisherigen Studienverläufe aller Studierenden die zugrundeliegende Datenbasis.

Das New-Item Problem ist hier problematischer. An dieser Stelle könnte zum Beispiel über händische Einträge oder eine Hybridisierung dem Effekt entgegengewirkt werden.

3.6 Hybridisierungsmöglichkeiten

In [Bu02] wurde dargelegt, dass eine kombinierte Lösung aus mehreren Ansätzen potentiell die Schwächen der einzelnen Ansätze beseitigen kann.

Der von uns dargestellte Ansatz ist in der Reinform zwar verwendbar, hilft aber vor

allem bei Modulen, welche nur von einzelnen Fachrichtungen belegt werden. Und auch dann nur, wenn bereits einige entsprechende Vorgängermodule bekannt sind, sodass das System auf diesen Modulen aufbauen kann.

Um eine präzisere Empfehlung generieren zu können, bietet sich eine Kombination mit anderen Ansätzen an. Im Folgenden werden einige beispielhafte Hybridisierungen kurz skizziert.

Bereits am Anfang dieses Kapitels wird erwähnt, dass es sinnvoll ist, nur die Module vorzuschlagen, welche laut der Studienordnung des Studierenden möglich sind. Diese Einschränkung ist ein Beispiel für die Kombination mit einem Knowledge-basierten Ansatz, da hier das Wissen um die Bedürfnisse des Studierenden herangezogen wird. Auch eine Berücksichtigung der Leistungspunkte, die ein Modul bringt bzw. ein Studierender in einem bestimmten Bereich noch benötigt, fällt in diesen Bereich.

Ein anderes Beispiel wäre die Hybridisierung mit dem demographischen Ansatz. So könnte beispielsweise bereits vor der Analyse auf Vorgänger oder Nachfolger die Menge der betrachteten Studierenden eingeschränkt werden. Verwendet man nur die Daten der Studierenden, die beispielsweise den gleichen Studiengang wie der User belegen und sich innerhalb der letzten 6 Jahre eingeschrieben haben, erhält man einen deutlich spezifischeren Überblick der Vorgänger und Nachfolger. Nimmt man andere Kriterien als „Studiengang“ und „Einschreibungszeitpunkt“, lassen sich so diverse Personengruppen eingrenzen, sodass der Benutzer ggf. eine auf ihn abgestimmte Teilgruppe aller Studierenden als Referenzgruppe erhalten würde. Dabei ist als Kriterium alles denkbar, was dem System über die User bekannt ist.

Um dem New-Item Problem entgegenzuwirken wäre eine Kombination mit einem gegen dieses Problem resistenten Ansatz, wie dem Content-basierten Ansatz denkbar. Hierbei wird die Ähnlichkeit der Items unter Berücksichtigung vorher definierter Eigenschaften der Items verwendet, um eine Empfehlung auszusprechen. Werden beide Ansätze kombiniert, erhält das Item dadurch eine Möglichkeit, empfohlen zu werden, und so selbst Vorgänger- und Nachfolgerbeziehungen aufzubauen.

4 Ausblick

Das in diesem Beitrag vorgestellte Konzept ist in dieser Form noch nicht implementiert. Eine Umsetzung wäre vor allem in Hinblick auf eine Evaluierung der Präzision des Ansatzes sinnvoll.

Auch eine genauere Betrachtung der möglichen Hybridisierungsmöglichkeiten ist wünschenswert. Dabei müsste evaluiert werden, welche Hybrid-Ansätze potentiell effektiv sind, also möglichst zutreffende Empfehlungen generieren. Anschließend müsste der genaue Effekt der Hybrid-Ansätze auf die Qualität der Empfehlungen verglichen und analysiert werden. Auch eine Kombination mehrerer Hybridansätze ist

dabei denkbar.

Außerdem ist die Entwicklung eines Sandbox-Systems denkbar. Hier könnte ein Studierender sehen, welche Auswirkungen eine Modulauswahl auf die weiteren Empfehlungen durch das System hat.

Literaturverzeichnis

[FH09] Fischer, H.; Hartau, C.: STiNE an der Universität Hamburg – Zur Einführung eines integrierten Campus-Management-Systems. 9. Internationale Tagung Wirtschafts-informatik

[Ri09] Rieger, B.; Haarmann, T.; Höckmann, E.; Lüttecke, S.: Data Warehouse basierte Entscheidungsunterstützung für das Campus-Management von Hochschulen. 9. Internationale Tagung Wirtschaftsinformatik

[Bu02] Burke, R. D.: Hybrid Recommender Systems: Survey and Experiment. User modeling and user-adapted interaction 12.4 (2002): S. 331-370.

[SK09] Xiaoyuan Su;Taghi M. Khoshgoftaar: A Survey of Collaborative Filtering Techniques. Advances in Artificial Intelligence, vol. 2009, Article ID 421425, 19 pages, 2009. doi:10.1155/2009/421425

[He04] Herlocker, J.L. et.al.: Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS) 22.1 (2004): 5-53.

[GT05] Adomavicius, G.; Tuzhilin, A.: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. Knowledge and Data Engineering, IEEE Transactions on 17.6 (2005): 734-749.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Effizientes und bedarfsgerechtes Berichtswesen auf Basis eines modularen Systems aus Open-Source-Produkten

Thomas Jankowski [1], Ulrike Lucke[2]

Abstract: Data-Warehouse-Systeme bzw. Business-Intelligence-Lösungen finden mittlerweile auch in der Hochschullandschaft eine immer weitere Verbreitung. Dabei kann auf fertige Lösungen zum Beispiel von Anbietern der Campus-Management-Systeme gesetzt werden, aber auch auf unabhängige Lösungen. Die Universität Potsdam hat sich für eine eigenständige Lösung auf Basis von flexiblen Open-Source-Produkten entschieden, um damit die vorhandenen Verwaltungsprozesse bestmöglich zu unterstützen. Erste Bereiche (Bewerber-, Studierenden- und Finanzverwaltung) wurden integriert. Der Beitrag gibt einen Überblick über die entwickelte Lösung und weist auf positive Erfahrungen, aber auch auf zu meisternde Herausforderungen hin.

Keywords: Business Intelligence, Data Warehouse, Berichtswesen, Hochschulstatistik, Systemarchitektur, Open Source

1 Einleitung und Begriffsklärung

Unter dem Schlagwort *Business Intelligence* (BI) haben rechnergestützte Formen der zielgerichteten Auswertung von Datenquellen eine zunehmende Bedeutung erlangt. An Hochschulen geht dies i.d.R. aus den gesetzlich vorgeschriebenen Pflichten (z.B. [BStatG87]) des *Berichtswesen* bzw. der *Statistik* hervor und betrifft v.a. die Bereiche Studium/Lehre und Forschung. Neben der hier seit langem bestehenden Herausforderung der Verarbeitung mit einer großen Vielzahl und Vielfalt von Datenquellen zeichnet sich Business Intelligence darüber hinaus durch das interaktive Arbeiten mit großen, komplexen Datenmengen aus und wird zunehmend als ein wichtiges Instrument der Hochschulsteuerung verstanden [Le04] (auch im Kontext der Systemakkreditierung).

In diesem Zusammenhang kommt dem Konzept eines sog. *Data Warehouse* (DWH, eine leistungsfähige Datenbank) eine zentrale Rolle zu. Hier werden Daten aus verschiedenen Quellen zusammengeführt. Der Begriff *Data Warehouse System* bezeichnet darüber hinaus die Gesamtheit aller Komponenten der unteren drei Ebenen (Datenerfassung, Datenhaltung und Datenbereitstellung). Dies wird in Abbildung 1 veranschaulicht. Auf dieser Basis bietet ein BI-System einen „IT-gestützten Zugriff auf Informationen sowie die IT-gestützte Analyse und Aufbereitung dieser Informationen. Ziel dieses Prozesses

[1] Universität Potsdam, Zentrale Einrichtung für Informationsverarbeitung und Kommunikation, Am Neuen Palais 10, 14469 Potsdam, thomas.jankowski@uni-potsdam.de
[2] Universität Potsdam, Institut für Informatik, A.-Bebel-Str. 89, 14482 Potsdam, ulrike.lucke@uni-potsdam.de

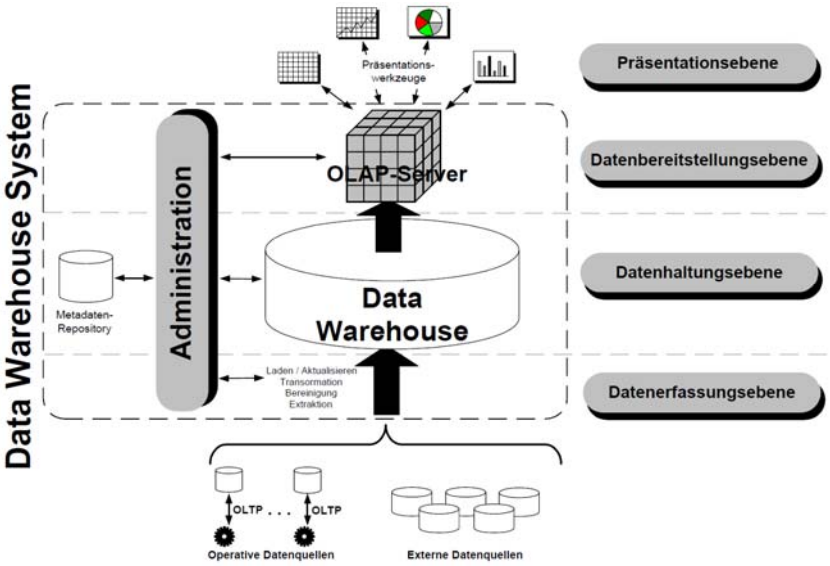


Abbildung 1: Komponenten eines Data Warehouse Systems nach [SBU99**]**

ist es, aus dem im Unternehmen vorhandenen Wissen neues Wissen zu generieren" [SG15]. Über die systematische Datenaufbereitung auf den unteren drei Ebenen hinaus gelangen insbesondere die anschauliche Präsentation von bzw. die gezielte Interaktion mit Daten in den Fokus. Zudem wird nicht nur von Daten, sondern von den daraus gewonnenen Informationen gesprochen. Weiterhin wird die Datenerfassung als Prozess verstanden (sog. ETL-Prozess: extract, tranform, load), für den es in der Praxis zusätzlicher organisatorischer Regelungen bedarf.

Mit Blick auf die IT-Infrastrukturen und Organisationsstrukturen sind bei der Einführung eines BI-Systems an Hochschulen zudem bestehende Rahmenbedingungen zu beachten. Dazu zählen v.a. die heterogenen IT-Landschaften mit oft mangelnden Schnittstellen und verteilten Verantwortlichkeiten. Das ergibt folgende Anforderungen an ein BI- System:

- flexible Schnittstellen zu vorhandenen Datenquellen (verschiedene Plattformen, Strukturen und Formate bis hin zu teilweise analoger Datenübermittlung)
- flexible Generierung von Statistiken (für externe und interne Anforderungen) auf Basis eines anpassbaren ETL-Prozesses
- ausführliche Metadaten (sowohl für die verwendeten Quelldaten als auch für die generierten Statistiken)
- dezentrale Nutzung & Bereitstellung von Berichten (sowohl in vordefinierter Form als auch ad hoc zusammengestellt)
- differenziertes Rechtekonzept für den verteilten Zugriff (in der Hoheit der für die Daten bzw. den Prozess fachlich Zuständigen)
- Protokollierung von Aktionen; Historisierung bestimmter Zustände der Datenbasis

Oftmals besteht zudem eine gewisse Angst vor komplexen Einführungsprojekten, auch angesichts schwergängiger Produkte [Me012][Ka13]. Dieses Hemmnis wird verstärkt durch ein i.d.R. geringes IT-Budget. Vor diesem Hintergrund beschreibt der Beitrag zunächst den aktuellen Stand der Technik, wobei der Fokus auf Open-Source-Produkte gelegt wird. Daraus werden eine Systemarchitektur und ein Verarbeitungsprozess abgeleitet und mit geeigneten Werkzeugen unterlegt. Abschließend werden das daraus resultierende System und die damit gemachten Erfahrungen beschrieben.

2 Stand der Technik

Die methodische Basis von Business Intelligence bildet das Online Analytical Processing (OLAP). Hier erfolgen interaktive Anfragen auf einem mehrdimensionalen Datenwürfel, der entlang struktureller und inhaltlicher Parameter gefiltert wird [Cla98]. Als praktische Umsetzung dieses Konzeptes haben sich zahlreiche Produkte auf dem Markt etabliert. Für alle vier o.g. Ebenen finden sich kommerzielle Lösungen von Spezialanbietern (z.B. HISinOne-BI[3] für Hochschulen oder qlikview[4] für Datenanalysen allgemein), aber auch von Allroundern (wie Microsoft Power BI[5], Oracle BI Foundation Suite[6], IBM Cognos BI[7] oder SAP Crystal Reports[8]). Dabei decken qlikview und Crystal Report nur die Präsentationsebene ab. Zudem ist eine Reihe von Open-Source-Werkzeugen verfügbar (wie z.B. Pentaho[9], JasperSoft[10] und RapidMiner[11]).

Die kommerziellen Werkzeuge sind in der für den Einsatzbereich Hochschule nötigen Konfiguration i.d.R. preisintensiv (Kaufpreis und Beratertage) – mit Ausnahme von HISinOne-BI, das aber stark auf die HIS-eigenen Datenquellen ausgerichtet ist. Mit Blick auf diese enge Integration in Standard-Software besteht hier aber noch einiges Potential. Die verfügbaren Open-Source-Werkzeuge halten jedoch einem Abgleich mit den Anforderungen der verschiedenen Nutzergruppen und den vorhandenen Datengrundlagen stand, auch weil sie kaum noch Unterschiede zu kommerziellen Lösungen bzgl. Wartungsaufwand zeigen. Tabelle 1 zeigt eine Bewertung der drei o.g. Werkzeuge anhand der in Abschnitt 1 identifizierten Anforderungen.

Dafür wurden die genannten Kriterien weiter ausdifferenziert[12] und jeweils die Eignung der Werkzeuge (Erfüllungsgrad in Stufen von 0/30/60/100%) geprüft. Gemäß den definierten Gewichten pro Kriterium/Gruppe entsteht die dargestellte Gesamtbewertung.

[3] https://www.his.de/produkte/hisinone/management/business-intelligence.html
[4] http://www.qlik.com/
[5] https://www.powerbi.com/
[6] http://www.oracle.com/us/solutions/business-analytics/
[7] http://www.ibm.com/software/products/de/business-intelligence/
[8] http://www.crystalreports.com/
[9] http://www.pentaho.de/
[10] https://www.jaspersoft.com/
[11] https://rapidminer.com/
[12] verkürzte Darstellung der ca. 50 einzelnen Kriterien; Details werden auf Anfrage bereitgestellt

Kriteriengruppe	Gewicht	Pentaho	JasperReports	RapidMiner
Systemschnittstellen	20,00%	100,00%	100,00%	100,00%
Einbindung vorhandener Datenquellen	15,00%	30,00%	30,00%	30,00%
ETL- und Statistik-Tools/ Benutzerfreundlichkeit	25,00%	75,00%	77,50%	69,50%
Kosten und Aufwand	15,00%	46,67%	46,67%	60,00%
Rollen und Rechte	15,00%	100,00%	100,00%	28,57%
Sonstiges (z.B. Support)	10,00%	80,00%	66,00%	80,00%
Gesamtbewertung		**73,25%**	**72,48%**	**63,16%**

Tabelle 1: Vergleich von Open-Source-Tools mit Anforderungen von Hochschulen

Im Ergebnis liegt Pentaho leicht vor JasperSoft, aber auch Rapidminer wäre zumindest eingeschränkt geeignet. Den Einsatz von Pentaho und JasperSoft im Rahmen eines modularen Systems beschreibt Abschnitt 3.

3 Technische Lösung

3.1 Systemkomponenten

Ausgehend von der vorgestellten Architektur eines DWH-Systems wurde mit Blick auf die Stärken und Schwächen der einsetzbaren Werkzeuge eine verteilte Lösung aus unterschiedlichen Komponenten konzipiert. Dabei werden die Datenbereitstellungs- und die Präsentationsschicht verschmolzen, da anstelle einer komplexen OLAP-Software die Abbildung der nötigen Datenstrukturen in einer relationalen Datenbank auf einfachere Weise möglich ist (sog. Relational OLAP [CMR02]). Die resultierende Architektur ist in Abbildung 2 dargestellt. Die einzelnen Ebenen werden nachfolgend kurz erläutert.

Operative Systeme

Hierzu zählen alle in der Hochschule eingesetzten Legacy-Tools für die Verwaltung zur Hochschulsteuerung relevanter Daten, wie für Studienbewerber, Studierende, Lehrveranstaltungen, Prüfungsleistungen, Personal, Finanzen, Drittmittel, Publikationen usw.

Datenerfassungsebene

Hier erfolgen die Extraktion, Transformation und das Laden der Daten in das DWH. Dabei werden komplexe ETL-Prozesse in kleine, handhabbare Teilprozesse zerlegt. Es kann auf beliebige Datenquellen zugegriffen werden (bspw. relationale Datenbanken, Excel-Dateien, strukturierte Textdateien). Für jede Datenquelle werden ein eigener ETL-Prozess und Importrhythmus definiert, die automatisch ausgeführt werden. In Abschnitt 3.2 sind die Möglichkeiten des ETL-Prozesses näher erläutert.

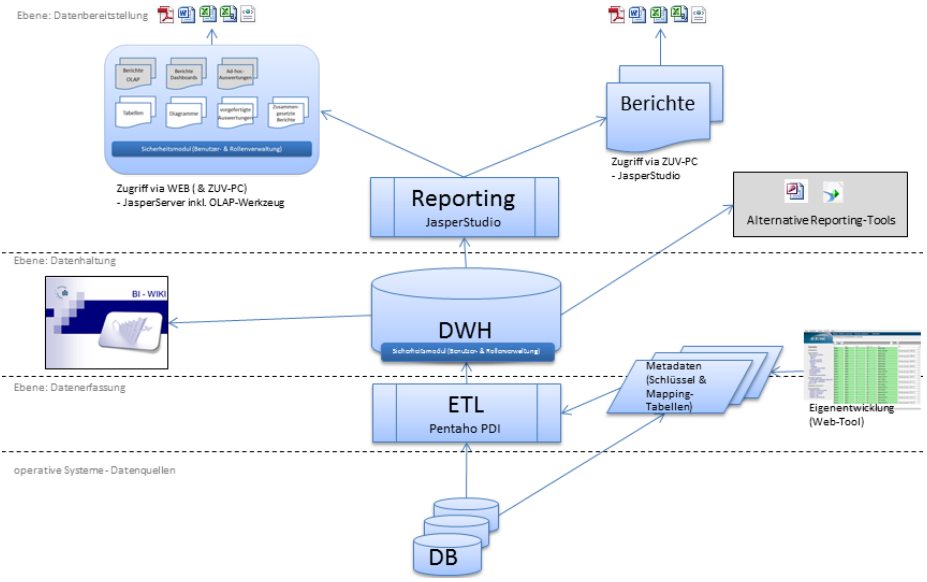


Abbildung 2: Architektur des modularen BI-Systems

Datenhaltungsebene

Zur Verwaltung der angebundenen Datenquellen werden zwei Tools eingesetzt.

- In einem Wiki werden beliebige Informationen zum Data-Warehouse-System manuell erfasst. Das sind insbesondere Erläuterungen zu den Berichten, Beschreibungen der Datenquellen pro Datenbereich, häufig gestellte Fragen und ihre Antworten, fachliche Ansprechpartner sowie Informationen zu den ETL-Prozessen. Das Wiki bietet die Möglichkeit, dass Informationen zu den Berichten dezentral durch die fachlich zuständigen Mitarbeiter gepflegt werden können.
- Zudem wird eine Web-Anwendung zur Datenverwaltung eingesetzt, die ebenfalls dezentral durch die fachlich zuständigen Mitarbeiter gepflegt werden kann. Das umfasst vier Arten von Daten: *Schlüsseltabellen* enthalten zusätzliche Angaben zu zu beliebigen Schlüsseln (bspw. Kostenstellenschlüssel), das können abweichende Gruppierungs- oder Sortierungsmerkmale oder auch überlagernde Bezeichnungen sein. *Mappingtabellen* enthalten Regeln zum Umschlüsseln von Daten, z.B. wenn in zwei operativen Systemen für die gleiche Kostenstelle unterschiedliche Schlüssel verwendet werden. *Technische Metadaten* enthalten bspw. Informationen über die erfolgten Datenimporte und erlauben den Fachanwendern die Festlegung, wann Daten als historisiert (d.h. unveränderbar) gekennzeichnet werden. Die Pflege *zusätzlicher fachlicher Daten* ist notwendig, wenn es für bestimmte Informationen kein zugrundeliegendes operatives System gibt, die Daten aber dennoch in Statistiken verwendet werden sollen. Dies kann auch sinnvoll sein, wenn sich aufgrund geringer Datenmengen die Entwicklung eines Importprozesses nicht lohnt.

Datenbereitstellungs- und Präsentationsebene

Es werden Werkzeuge eingesetzt, welche OLAP-Analysen und die Erstellung von Standardberichten auf Basis der relationalen Datenbank des Data-Warehouses ermöglichen. Die Inhalte der Standardberichte werden im Moment der Berichtsanforderung aktuell ermittelt und können durch den Anwender dynamisch über Filterkriterien eingeschränkt werden.

3.2 Prozess von der Datenerhebung bis zur Statistik

Neben der technischen Lösung bedarf es auch organisatorischer Vereinbarungen für die Bereitstellung und Verarbeitung von Daten, die wiederum technisch abgebildet werden müssen. Abbildung 3 zeigt die dafür nötigen Schritte und die beteiligten Akteure; dies sind i.d.R. die Fachabteilung und der Data-Warehouse-Experte.

Es sind acht Schritte vom Erkennen des Statistikbedarfs bis zur Veröffentlichung einer Statistik erkennbar. Zunächst muss die Fachabteilung in der Lage sein, die gewünschten Statistiken/Berichte und die anzuzeigenden Merkmale (Daten) klar zu beschreiben (Schritt 1). Anhand dieser Informationen kann der DWH-Experte erkennen, welche Daten zusätzlich benötigt werden (Schritt 2). Anschließend beginnt eine detaillierte Analyse, welche Daten tatsächlich vorhanden sind, welche zusätzlich erhoben werden müssen und welche nicht bereitgestellt werden können (Schritt 3). In der Folge kann die Datenbankstruktur im DWH entworfen und realisiert werden (Schritt 4). Dann folgt mit der Realisierung des ETL-Prozesses der aufwendigste Teil (Schritt 5). Hierbei geht es nicht nur darum, die Daten von einer Datenquelle in das DWH zu kopieren. Vielmehr können diverse Verarbeitungsschritte in dem ETL-Prozess integriert werden:

- Plausibilitätsprüfung (Sind alle Pflichtangaben vorhanden? Sind die Schlüssel gültig? Gibt es Verstöße gegen inhaltliche oder fachliche Regeln?)
- automatische Fehlerkorrekturen (soweit möglich)
- Schlüssel-Mapping (Zuordnung von Schlüsselwerten nach definierten Regeln)
- Prozesskonstanten und -variablen festlegen (z.B. Zeitstempel als Importdatum)
- dynamische Wahl des ETL-Prozess-Pfades in Abhängigkeit von den zu importierenden Daten (Beispiel: Drittmittelfinanzdaten durchlaufen einen anderen ETL-Pfad, als Haushaltsmittelfinanzdaten. Somit können andere Plausibilitäten abgebildet werden, auch wenn es sich um die gleiche Schnittstelle handelt.)
- Gruppierung und Aggregation von Daten
- Ergänzung neuer Merkmale nach festgelegten Regeln
- Fehlerbehandlung (d.h. Erzeugung und Versand von Fehlerprotokollen, ggf. Löschung der Daten des letzten Imports, Benachrichtigung des fachlich Zuständigen, Abbruch des Importvorgangs)

Im Schritt 6 liegen die notwendigen Daten bereits im DWH vor. Anschließend müssen die organisatorischen Rahmenbedingungen für die Nutzung der Daten geschaffen werden (Schritt 7), bevor die Berichte und Statistiken freigeschaltet werden (Schritt 8).

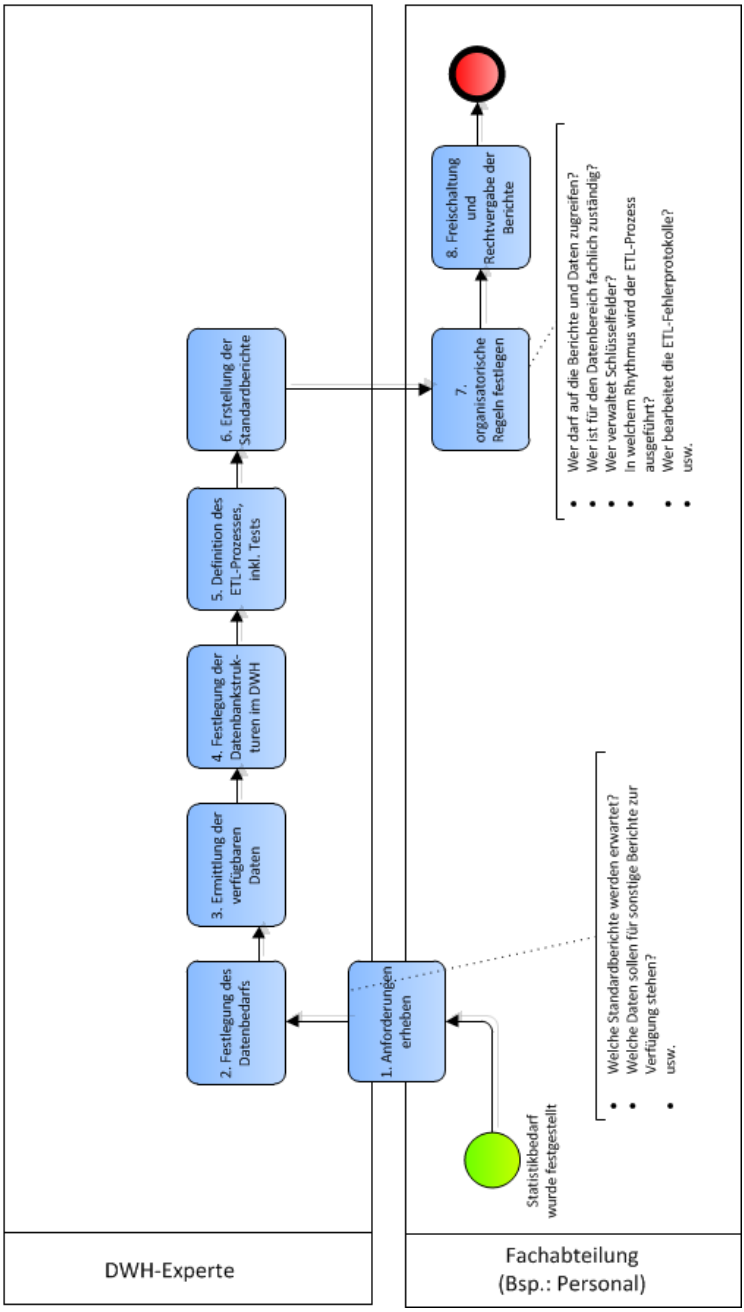


Abbildung 3: DWH-Gesamtprozess – von der Datenerhebung bis zur Statistik

3.3 Eingebundene Open-Source-Werkzeuge

In dem in Abschnitt 2 vorgestellten Produktvergleich liegt Pentaho insgesamt leicht vor JasperSoft. Hinter diesen beiden Produktnamen verbergen sich jedoch eine Sammlung von Open-Source-Tools, die jeweils für einen bestimmten Einsatzbereich (bspw. Datenerfassung oder Präsentation) gedacht sind. Diese sind teilweise aufeinander abgestimmt, können aber auch separat genutzt werden. Schon während der Produktbewertung zeigte sich, dass die Community-Version von Pentaho zwar geringfügig bessere Funktionen bietet, aber im Präsentationsbereich vereinzelt Probleme auftraten (verschwindende Felder, unerklärliche Fehlermeldungen).

Gemäß der entwickelten Systemarchitektur fiel die Entscheidung daher differenziert aus. Für die Ebene der Datenerfassung wurde zur Definition der ETL-Prozesse Pentaho gewählt. Pentaho hatte im Gesamtvergleich die höchste Bewertung erhalten, und ließ sich im ETL-Bereich auch problemlos ohne Support durch den Hersteller implementieren. Die Datenbereitstellung und die Präsentation (sowohl von OLAP-Analysen als auch von Standard-Berichten) erfolgen jedoch mit Hilfe von Jaspersoft. Jaspersoft schnitt im Gesamtvergleich nur geringfügig schlechter als Pentaho ab, konnte jedoch auch ohne Support fehlerfrei implementiert werden. Für die Ebene der Datenhaltung war keine neue Lösung notwendig; hier wurde der vorhandene Postgres-Datenbank-Cluster der Universität genutzt.

4 Beispielhafte Umsetzung

Nachfolgend wird die Umsetzung für Statistiken aus dem Bereich der Hochschulbewerberdaten (Studienbewerbung und Immatrikulation) dargestellt – von konkreten Datenquellen über deren Import / Verarbeitung hin zu generierten Statistiken und deren abgestufter Freigabe für verschiedene Nutzergruppen. Andere Domänen wie z.B. Forschungsinformationssysteme können auf vergleichbare Weise gestaltet bzw. angebunden werden.

4.1 ETL-Prozess zur Datenerfassung

Die Definition des Importprozesses erfolgt in Pentaho grafisch. Dabei können ETL-Prozesse in beliebige Sub-Prozesse gegliedert und grafisch modelliert werden. Prozesse können zu Analysezwecken schrittweise durchlaufen und der Datenzustand bei jedem Prozessschritt geprüft werden. Abbildung 4 zeigt den Hauptprozess für Bewerberdaten, der als Datenquelle HISinOne (Modul APP) nutzt. Jeder Prozess hat ein Start- und ein Endereignis und gliedert sich in diverse Schritte. Dabei kann es sich um Einzelarbeitsschritte (grüne Symbole) oder um komplexe Sub-Prozesse (orange Symbole) handeln. Die Prozessschritte können nacheinander oder auch parallel ablaufen. Bei parallel ablaufenden Prozessschritten gibt es zuvor eine Weiche, welche bestimmte Bedingungen abprüft, wie in Abbildung 4 dargestellt. Die Wahl des Weges wird jedoch

in diesem Beispiel nicht automatisiert durch eine Bedingung gesteuert, sondern der Prozessverantwortliche hat hier die Möglichkeit manuell einzugreifen und den linken (regulärer, d.h. täglicher Datenimport) oder den rechten Pfad (initialer, d.h. einmaliger Datenimport) festzulegen.

Jeder Importschritt kann wiederum aus einem Sub-Prozess bestehen. Der Prozess kann schrittweise durchlaufen werden, wobei nach jedem Schritt die Daten geprüft werden können. Im Laufe des ETL-Prozesses können beliebige Protokolle geschrieben werden, entweder in die Datenbank oder in das File-System. Es ist auch möglich, in Abhängigkeit von bestimmten Bedingungen (z.B. beim Auftreten eines Fehlers) E-Mail-Benachrichtigungen zu versenden.

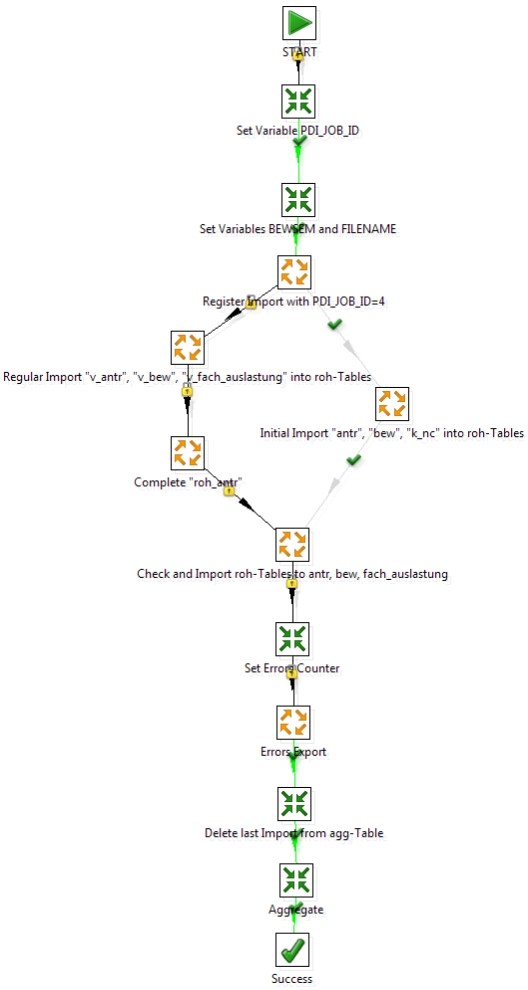


Abbildung 4: Grafische Modellierung des ETL-Prozesses in Pentaho

4.2 Berichtsdefinition und -bereitstellung

Die Berichte und Statistiken werden mit JasperStudio erstellt, wie Abbildung 5 anhand einer Auswertung von Bewerberdaten zeigt. Neben den primären Ausgabeelementen (vordefinierte Texte und Grafiken; dynamisch erzeugte Tabellen, Diagramme, Berichte, Grafiken, Verzeichnisse usw.) können auch Steuerungselemente (z.B. Selektions- und Sortierparameter) und eigene Java-Script-Bausteine spezifiziert werden.

Die Berichte können über Jasperserver im Web angeboten und mit Rechtedefinitionen versehen werden. Im Ergebnis können sich Angehörige der Hochschule, gemäß den für sie definierten Freigaben, Berichte online ansehen bzw. analysieren. Die Berichte werden in einer Baumstruktur dargestellt, wie in Abbildung 6 dargestellt ist. Zusätzlich

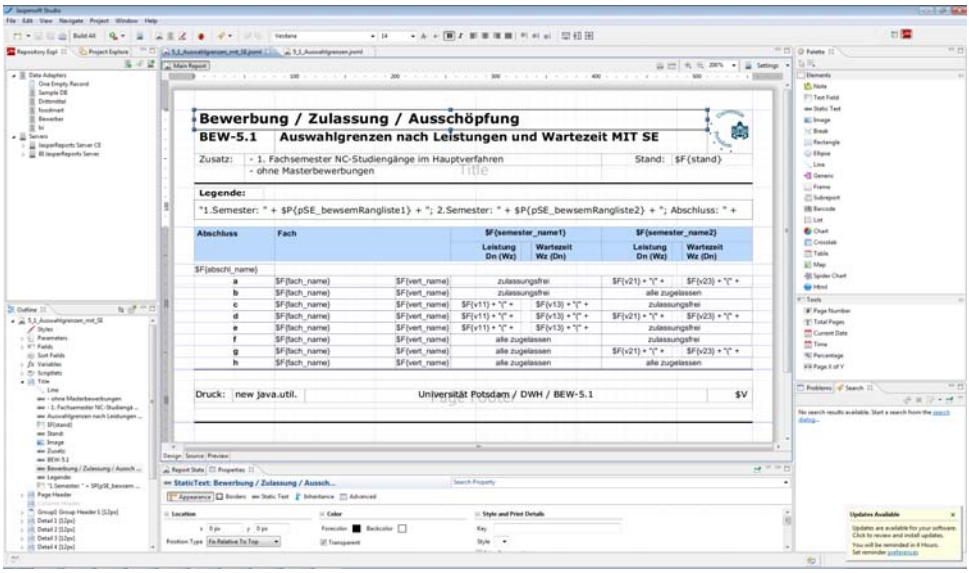


Abbildung 5: Definition einer Statistik mittels JasperStudio

kann nach Stichworten gesucht oder in Drittanwendungen (PDF, Excel usw.) exportiert werden. Vor der Ausführung eines Berichts ist es möglich, Filterkriterien (Parameter) vom Nutzer abzufragen. Zudem können Daten mit Hilfe eines OLAP-Werkzeuges analysiert werden; dies ist hier jedoch nicht dargestellt.

5 Zusammenfassung und Ausblick

In Kontrast zu den etablierten BI-Lösungen „aus einem Guss“ wird in diesem Beitrag ein modulares System für die Erstellung von Berichten und Statistiken ausgehend von den in vorhandenen Werkzeugen vorgehaltenen Datenbeständen propagiert. Der beschriebene Ansatz ist an der Universität Potsdam für das Themengebiet Studierendendaten bereits umgesetzt; weitere Gebiete werden aktuell integriert. Der dargestellte Weg erfordert internes Personal, welches sich mit den Fachdaten, der Erstellung der Datenbanken, der Entwicklung von Datenbankabfragen und dem Berichtsentwurf auseinander setzt. Dabei ist die Implementierung der ersten Schnittstelle (Datenquelle) zeitlich am aufwendigsten. Die Anbindung der ersten Datenquelle bis hin zur Entwicklung von zwölf Statistiken dauerte effektiv etwa 90 Personentage.

Bereits jetzt sind Auswirkungen auf Entscheidungsprozesse der Hochschule erkennbar. So werden bspw. Bewerbungs- und Zulassungsverfahren durch tagesaktuelle Berichte bereits während der Bewerbungsphase unterstützt, und die Hochschulplanung profitiert

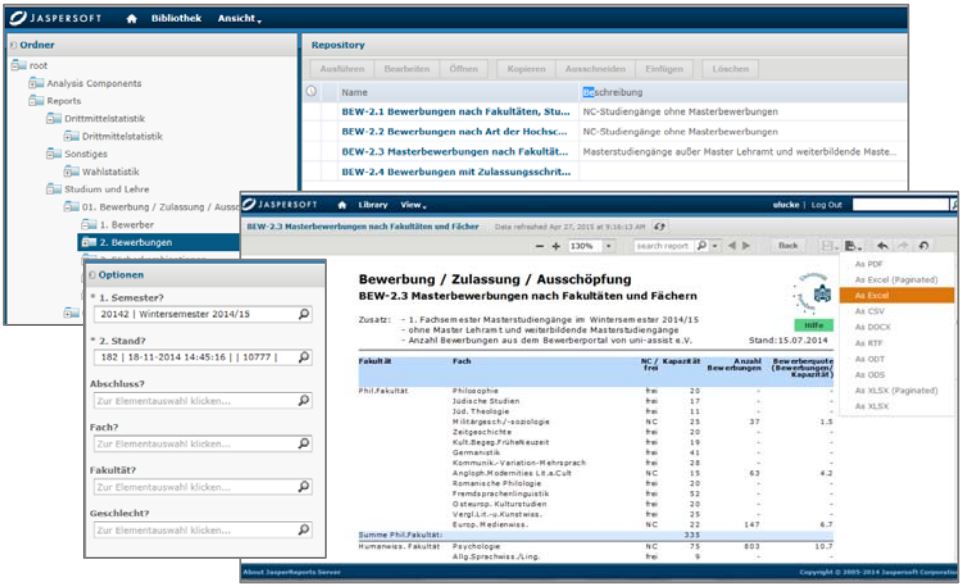


Abbildung 6: Zugriff auf die verfügbaren Berichte in JasperServer

u.a. von detaillierten Auslastungsberichten der Studiengänge. Im Vergleich zu dem früher sehr aufwändigen, langwierigen und fehleranfälligen Verfahren der Erhebung und Auswertung statistischer Daten ist hier nicht nur eine klare Verbesserung für die zentrale Verwaltung bzw. die Hochschulleitung erkennbar, sondern auch die dezentralen Prozesse in den Fakultäten und Einrichtungen der Hochschule werden vereinfacht. Das verbessert die wahrgenommene Datenqualität und so den zweckmäßigen Einsatz [Vo14]. Zugleich wird die Zugänglichkeit der Datenbestände erhöht, was die Transparenz von Entscheidungsprozessen und damit allgemeine Werte wie Identifikation, Beteiligung und Veränderungsbereitschaft signifikant fördert.

Auch aus technischer Perspektive ist die gewählte Lösung positiv zu beurteilen. Die für das Einführungsprojekt vorgesehene Mittelplanung wurde unterschritten: Zum Einen konnte der Sachkostenansatz auf die Beschaffung von Hardware reduziert werden. Darüber hinaus hat sich auch im Bereich Personalkosten eine Entlastung durch die Verlagerung von Aufwänden ergeben, die sich aber aufgrund der noch nicht reibungsfreien Abläufe bei der Datenaufbereitung derzeit nicht verlässlich quantifizieren lassen: Einerseits werden in den Fachabteilungen doppelte Arbeiten beim Sammeln und Transformieren von Daten vermieden, andererseits traten gerade in der Einführungsphase vorübergehende Mehrbelastungen auf. Die unmittelbare Implementierung der lokalen Prozesse und Datenstrukturen war jedoch weniger komplex als die Einführung eines vglw. starren Komplettsystems. Das Risiko des weiteren Betriebs der entwickelten Lösung konnte durch den Rückgriff auf standardisierte Architekturmuster und Support-Verträge aufgefangen werden. Zudem konnte durch den modularen Ansatz und die

Nutzung von open-source-Produkten die Herstellerabhängigkeit reduziert werden. Die dadurch gewonnenen Freiheitsgrade werden als wichtiger eingestuft als eine etwaige Komplettunterstützung nach den Vorgaben eines Anbieters. Als besonders empfehlenswert hat es sich erwiesen, die vorhandenen Datenquellen sukzessive anzubinden und den ETL-Prozess pro Datenquelle in Sub-Prozesse zu gliedern. So wird die Komplexität der Einführung stark reduziert, was insbesondere die Fachabteilungen (neben dem regulären Tagesgeschäft) entlastet.

Die in diesem Beitrag beschriebene Vorgehensweise ist insbesondere für Hochschulen mit einer starken Entwicklungsabteilung geeignet, sowie für Domänen die eine Separierung in einzelne Prozesse und Datenstrukturen erlauben.

Literaturverzeichnis

[BStatG87] Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG), 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 13 des Gesetzes vom 25. Juli 2013 (BGBl. I S. 2749). http://www.gesetze-im-internet.de/bstatg_1987/

[Cla98] Clausen, N.: OLAP – Multidimensionale Datenbanken. Addison-Wesley-Longman, Bonn 1998.

[CMR02] Colossi, N.; Malloy, W.; Reinwald, B.: Relational extensions for OLAP, IBM Systems Journal 41(4), 2002, S. 714 - 731.

[Ka13] Kaiser, S.; Kuhnt, E.; Lemcke, S.; Lucke, U.: Web-basierte Beschaffung. In Proc. INFORMATIK 2013, LNI P-220, Bonn : Köllen, 2013, S. 308-319.

[KR13] Kimball, R.; Ross, M.: The Data Warehouse Toolkit – The Definitive Guide to Dimensional Modeling (3. Auflage), Wiley 2013.

[Le04] Leszczensky, M.; Orr, D.; Schwarzenberger, A.; Weitz, B.: Staatliche Hochschulsteuerung durch Budgetierung und Qualitätssicherung, in: Hochschulplanung (167), HIS Hochschulinformationssysteme GmbH, Hannover 2004.

[Me12] Mertens, P.: Schwierigkeiten mit IT-Projekten der öffentlichen Verwaltung. Informatik-Spektrum, 35/06, S. 433-446, 2012.

[SG15] Springer Gabler Verlag (Herausgeber); Gabler Wirtschaftslexikon;,Stichwort: Business Intelligence. http://wirtschaftslexikon.gabler.de/Archiv/75968/business-intelligence-v10.html

[SBU99] Sinz, E.J.; Böhnlein, M.; Ulbrich-vom Ende, A.: Konzeption eines Data Warehouse-Systems für Hochschulen, in: Proc. Informatik '99 - Informatik überwindet Grenzen, 29. Jahrestagung der Gesellschaft für Informatik, Springer 1999, S. 111-124.

[Vo14] Vogel, D.; Löbel, St.; Proeller, I.; Schuppan, T.: Einflussfaktoren von Führungsverhalten in der öffentlichen Verwaltung. der moderne staat, 07/02, 2014, S. 459-478.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Workshop „Strategisches Informationsmanagement und IT-Governance im Öffentlichen Sektor“ (SIMÖS)

Workshop „Strategisches Informationsmanagement und IT-Governance im Öffentlichen Sektor“ (SIMÖS)

Abstract zum Workshop
Der Workshop beschäftigt sich mit dem Thema Strategisches Informationsmanagement und IT-Governance im Öffentlichen Sektor. Darunter werden die Branchen Öffentliche Verwaltung, Gesundheitswesen und Non-Profit-Unternehmen des Öffentlichen Sektors gezählt.
Dabei resultierten nach einem doppelblinden Begutachtungsverfahren vier ausgewählte Beiträge. Zudem wurde ein Keynote-Vortrag zu IT4IT eingeladen. Dabei handelt es sich um eine Standardisierungsgruppe der OpenGroup zum Thema IT-Management und semantische Integration, Architektur und Standardisierung (generische Domänenontologie zur IT).
Der Betrag von Schaaf et al. thematisiert die Entwicklung und den Einsatz einer Domänenontologie des Informationsmanagements im Krankenhaus. Die Ontologie wird als Grundlage für die Entwicklung eines entscheidungsunterstützenden Werkzeugs und einer Variante zur Unterstützung der Lehre des Informationsmanagements im Krankenhaus genutzt.
Der Betrag von Kücherer et al. thematisiert vor dem Hintergrund, dass diverse Krankenhäuser heute bereits auf ITIL-Prozesse zur Unterstützung des IT-Servicemanagements setzen, die qualitative Analyse auf Basis der Informationsbedarfe und des Informationsmanagements im Krankenhauses die Adoption und das Managements der IT auf Basis von ITIL.
Der erste Betrag von Walser et al. untersucht auf Basis von früheren quantitativen Studien die Mechanismen des Performance Managements in Schweizer Krankenhäusern. Dies nimmt insbesondere vor dem Hintergrund immer größere Bedeutung ein, da die Schweizerischen Krankenhäuser in den letzten Jahren Diagnosis Related Groups (DRG’s) und eine neue Krankenhausfinanzierung eingeführt haben und damit der IT-unterstützten Steuerung der Performance von Krankenhäusern vor dem Hintergrund eines zunehmenden Wettbewerbs unter den Häusern eine zunehmende Bedeutung zukommt.

Der Betrag zweite Beitrag von Walser et al. geht der Frage des Business-IT-Alignments in Kommunen nach. Ebenfalls basierend auf früheren quantitativen Untersuchungen wird dabei der Frage nachgegangen, welche Maturität von Schweizer Gemeinden im Bereich Business-IT-Alignment haben. Dazu wurden mittels qualitativer Fallstudienforschung drei größere Kommunen der Schweiz untersucht und die daraus resultierenden Ergebnisse präsentiert. Im Rahmen des Beitrags wird auch ein einfaches Maturitätsmodell für die Analyse und Weiterentwicklung des Business-IT-Alignments in Gemeinden präsentiert.
Vor dem Hintergrund der Tatsache, dass im Gegensatz etwa zur Privatwirtschaft die Maturität des IT-Managements je nach Ausprägung teilweise gering ist, werden in Workshop-Form zuhanden der GI-Fachgruppe Verwaltungsinformatik Empfehlungen entwickelt, wie das strategische Informationsmanagement, die IT-Governance, das IT-Servicemanagement sowie das Management von Unternehmensarchitekturen in der öffentlichen Verwaltung mehr Relevanz erhalten, was Förderungs- und Hinderungsfaktoren dafür sind sowie welche Maßnahmen zu ergreifen sind, um ganz generelle der Strategischen IT-Führung mehr Bedeutung zuzumessen. Für die öffentliche Verwaltung gilt heute analog der Privatwirtschaft: Ohne IT läuft nichts mehr. Im Gegensatz zur Privatwirtschaft scheint diese Einsicht aber noch wenig etwa in Form von CIO's in Geschäftsleitungen der öffentlichen Verwaltung Anwendung gefunden zu haben. Ebenfalls teilweise unterentwickelt sind Einsätze von Rahmenwerken wie COBIT und ITIL sowie Standards in ähnlichen Bereichen.

Organisationskomitee des Workshops

Prof. Dr. Konrad Walser
Professor für Wirtschaftsinformatik und E-Government

Berner Fachhochschule
E-Government Institut
Brückenstrasse 73
CH-3005 Bern

E-Mail: konrad.walser@bfh.ch

Telefon +41-79-648 21 33

Prof. Dr. Matthias Goeken
Dozent für Wirtschaftsinformatik und Volkswirtschaftslehre

Hochschule der Deutschen Bundesbank
Schloss
D-57627 Hachenburg

E-Mail:
matthias.goeken@bundesbank.de
Telefon: +49-178-563-44-33

Programmkomitee des Workshops

Das Programmkomitee besteht aus folgenden interdisziplinär (Informatik, Medizininformatik, Versorgungsforschung, Verwaltung …) ausgewählten Persönlichkeiten aus Hochschulen und Praxis:

- Prof. Dr. Can Adam Albayrak – Hochschule Harz, Wernigerode
- Prof. Roland Blomer – Institute for Health Information Systems, Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik GmbH (UMIT), Tirol
- Prof. Dr. Michael Breidung – Eigenbetrieb IT und Org., Landeshauptstadt Dresden
- Prof. Dr. Andreas Breiter – IFIB, Universität Bremen
- Julian Detzel – CapGemini, Berlin
- Jochen Dreß – Zentrum für klinische Studien, Köln
- Georg Gelhausen – Bundesministerium für Inneres, Bonn
- Prof. Dr. Oliver Grieble – Hochschule Neu-Ulm
- Prof. Dr. Georg Rainer Hofmann – Hochschule Aschaffenburg
- Prof. Dr. Jürgen Holm – Berner Fachhochschule, Technik und Informatik
- Franziska Jahn – Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig, Leipzig
- Prof. Dr. Peter Klutke – Fakultät für Informatik, Hochschule Kempten
- Prof. Dr. Matthias Knoll – Hochschule Darmstadt
- Dr. Mike Krey – ZHAW, Winterthur
- Prof. Dr. Dagmar Lück-Schneider – HWR Berlin
- Prof. Dr. Tino Schuppan – IFG.CC Potsdam
- Prof. Dr. Anne-Dore Uthe – Hochschule Harz, Halberstadt
- Jörn von Lucke – Zeppelin-Universität, Friedrichshafen
- Dr. Martin Wind – Abteilungsleiter, Senatorin für Soziales, Kinder, Jugend und Frauen, Stadt Bremen
- Dr. Till Winkler, Departement of Management, Copenhagen Business School
- Prof. Dr. Alfred Winter, Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig, Leipzig; Leiter der GMDS-Arbeitsgruppe „Methoden und Werkzeuge für das Management von Krankenhausinformationssystemen“ der GMDS
- Dr. Petra Wolf – Fakultät für Informatik, Technische Universität München

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Entwicklung und Einsatz einer Domänenontologie des Informationsmanagements im Krankenhaus

Michael Schaaf[1], Franziska Jahn[2], Kais Tahar[3], Christian Kücherer[4], Alfred Winter[5] und Barbara Paech[6]

Abstract: Beim Studium der Literatur zum Informationsmanagement sieht man sich auf Grund verschiedener theoretischer Ansätze und praxisorientierter Frameworks mit einer enormen Begriffsvielfalt konfrontiert. Insbesondere im Krankenhaus sind CIO und die Geschäftsleitung auf vernetzte Informationen angewiesen, um die Krankenhausziele effektiv unterstützen zu können. Die Entwicklung entscheidungsunterstützender Werkzeuge für CIOs und die Geschäftsleitung wird auf Grund der fehlenden Formalisierung der komplexen Zusammenhänge der Konzepte im strategischen Informationsmanagement, im IT Service Management oder der IT Governance unzureichend unterstützt. Der strategische Einsatz von Informationssystemkomponenten im Krankenhaus stellt auf Grund sinkender IT-Budgets eine große Herausforderung dar. In dieser Arbeit werden die Entwicklung und der Einsatz einer Domänenontologie des Informationsmanagements im Krankenhaus skizziert. Die Ontologie soll als Grundlage für die Entwicklung eines entscheidungsunterstützenden Werkzeugs und einer Variante zur Unterstützung der Lehre des Informationsmanagements im Krankenhaus genutzt werden.

Keywords: Semantisches Netz, Ontologieentwicklung, Ontologie, Informationsmanagement, Krankenhaus

1 Einleitung und Fragestellung

Das Informationsmanagement (IM) im Krankenhaus ermöglicht die effektive Unterstützung der Krankenhausziele durch sein Informationssystem (IS), erfordert im Gegenzug jedoch dessen systematische Planung, Steuerung und Überwachung. Das berufliche Umfeld eines CIOs (Chief Information Officer) in deutschen Krankenhäusern ist geprägt durch niedrige IT-Budgets, die IT-Kosten pro Mitarbeiter liegen ca. 30%, unter dem Median anderer untersuchter Branchen [MSM08]. Niedrige IT-Budgets

[1] Universität Leipzig, Institut für Medizinische Informatik, Statistik und Epidemiologie, Härtelstraße 16-18, 04107, Leipzig, michael.schaaf@imise.uni-leipzig.de
[2] Universität Leipzig, Institut für Medizinische Informatik, Statistik und Epidemiologie, Härtelstraße 16-18, 04107, Leipzig, franziska.jahn@imise.uni-leipzig.de
[3] Universität Leipzig, Institut für Medizinische Informatik, Statistik und Epidemiologie, Härtelstraße 16-18, 04107, Leipzig, kais.tahar@imise.uni-leipzig.de
[4] Universität Heidelberg, Institut für Informatik, Lehrstuhl Software Engineering, Im Neuenheimer Feld 326, 69120 Heidelberg, christian.kuecherer@informatik.uni-heidelberg.de
[5] Universität Leipzig, Institut für Medizinische Informatik, Statistik und Epidemiologie, Härtelstraße 16-18, 04107, Leipzig, kais.tahar@imise.uni-leipzig.de
[6] Universität Heidelberg, Institut für Informatik, Lehrstuhl Software Engineering, Im Neuenheimer Feld 326, 69120 Heidelberg, paech@informatik.uni-heidelberg.de

spiegeln sich ebenfalls in geringen Mitarbeiterzahlen wider. Dennoch bestehen auch in Krankenhäusern Anforderungen an ein systematisches Informationsmanagement wie die Umsetzung eines IT Service Managements nach De-Facto Standards wie COBIT und ITIL und die Ausrichtung der Ziele des Informationsmanagements an den Krankenhauszielen [JTP11]. Hierfür benötigen sowohl CIO als auch die Krankenhausleitung ein tiefgreifendes Verständnis, wie das IS zum Erreichen der strategischen Krankenhausziele eingesetzt werden kann und eingesetzt wird. Eine große Herausforderung stellt daher die systematische und verständliche Bereitstellung der benötigten Informationen zur Erfüllung der IM-Aufgaben dar. Die Bereitstellung rein deskriptiver Kennzahlen zu Kosten und Effizienz des IS kann für die Beurteilung der komplexen Zusammenhänge nicht ausreichen, da Kennzahlen nicht alle entscheidungsrelevanten Informationen abbilden. Die langfristige Planung der Entwicklung im strategischen IM, die Bereitstellung von Informationen über die Projekte im taktischen IM und der Betrieb des IS im operativen IM erfordert die Bereitstellung vernetzter Informationen sowohl für den CIO als auch für die Krankenhausleitung. Darüber hinaus ist die Vernetzung der Informationen zum IM mit den durch das IS unterstützten Geschäftsprozessen sowie mit der Unternehmensstrategie erforderlich.

Diese komplexen Zusammenhänge müssen auch Studierenden der (Medizinischen) Informatik vermittelt werden, die in Vorlesungen zum Management von Informationssystemen im Gesundheitswesen auf spätere Leitungsfunktionen in diesem Bereich vorbereitet werden sollen. Für den Lehrenden und die Studierenden erfordert dies jedoch die sorgfältige Auseinandersetzung mit einer Begriffsvielfalt, verursacht durch unterschiedlichen problem-, aufgaben- und prozessorientierte Ansätze des Informationsmanagements [Kr09] und Frameworks wie COBIT und ITIL. Zahlreiche synonyme und homonyme Begriffe erschweren das Verständnis des Fachgebiets. Gleichzeitig fehlt es den verschiedenen Ansätzen und Frameworks an hinreichender Formalisierung, wodurch bedeutende Zusammenhänge zwischen strategischem, taktischem und operativem Management, IT Service Management und IT Governance unzureichend beschrieben und verstanden sind.

Das Ziel dieser Arbeit im Rahmen des DFG-Projekts „SNIK“ (Ein Semantisches Netz des Informationsmanagements im Krankenhaus) ist die Entwicklung einer Ontologie des Informationsmanagements im Krankenhaus, auf Basis des Lehrbuchs „Health Information System – Architectures and Strategies“ von Winter et al. [Wi11]. Es beschreibt einen problemorientierten Ansatz folgend, wie komplexe Krankenhausinformationssysteme hinsichtlich ihrer strategischen Planungs-, Steuerungs- und Überwachungsaufgaben systematisch gemanagt werden können. Dies wird – informal – mit prozessorientierten Frameworks (s.o.) verknüpft. Die entstehende Wissensbasis in Form einer Ontologie soll sowohl die Lehre unterstützen als auch in einem späteren Stadium die Grundlage für die Entwicklung eines entscheidungsunterstützenden Werkzeugs für den CIO bilden.

Im Bereich der Lehre soll die Ontologie für die Erarbeitung neuartiger Lehrkonzepte zur

Vermittlung einer einheitlichen Terminologie des Informationsmanagements im Krankenhaus genutzt werden. Die visuelle Repräsentation derartig vernetzter Informationen durch Ontologien ist in besonderer Weise geeignet, da sich einzelne Teilbereiche nachvollziehbar in den Gesamtkontext der Domäne einordnen lassen. Darüber hinaus können Ontologien in der Lehre durch Formulierung formalisierter Anfragen (z.B. mit SPARQL) für die Prüfungsvorbereitungen der Studenten genutzt werden.

Ein entscheidungsunterstützendes Werkzeug für den CIO, das auf den in der Ontologie dargestellten Zusammenhängen des Informationsmanagements basiert, kann die Navigation durch die vernetzten Informationen des Informationsmanagements unterstützen. Es könnte entscheidungsrelevante Informationen aus unterschiedlichen Quellen zusammenführen und bei der Erstellung von Berichten für die Krankenhausleitung unterstützen. Die Ontologie soll hierfür das Datenmodell für die zukünftige Umsetzung in einem entscheidungsunterstützenden Werkzeug bereitstellen.

In Kapitel 2 wird das methodische Vorgehen bei der Entwicklung der Ontologie des Informationsmanagements im Krankenhaus beschrieben. Im Ergebniskapitel 3 wird die Ontologie quantitativ beschrieben sowie in Ausschnitten präsentiert. Kapitel 4 diskutiert die Verwendung der Ontologie in der Lehre und gibt einen Ausblick in zukünftig geplante Entwicklungsschritte zu deren Erweiterung.

2 Methoden

2.1 Ontologieentwicklung

Für die formale Beschreibung und Repräsentation von Wissen einer Domäne steht das Konzept der Ontologie zur Verfügung. Die grundlegende Idee von Ontologien ist die Beschreibung einer Terminologie, indem die verwendeten Begriffe und Beziehungen formal festgehalten werden. Das Ziel der Formalisierung beschreiben Gómez-Pérez, Fernández-López und Corcho [GFC04] wie folgt: „...we can say that ontologies aim to capture consensual knowledge in a generic way, and that they may be reused and shared across software applications and by groups of people“. Die Entwicklung einer Ontologie des Lehrbuchs „Health Information Systems – Architectures and Strategies“ wurde in Anlehnung an den bestehenden Ansätzen „Ontology Development 101“ [NM01] und „Methontology“ [FGJ97] durchgeführt, wobei das konkrete Vorgehen der einzelnen Teilschritte zunächst spezifiziert wurde. Die Spezifikation wurde ausführlich in [Ja14] beschrieben. Wie die einzelnen Schritte in Abbildung 1 zeigen, handelt es sich bei der Ontologieentwicklung um einen iterativen Prozess, der zur kontinuierlichen Verbesserung der Qualität der Ontologie auch Rücksprünge zu vorangegangenen Schritten beinhaltet.

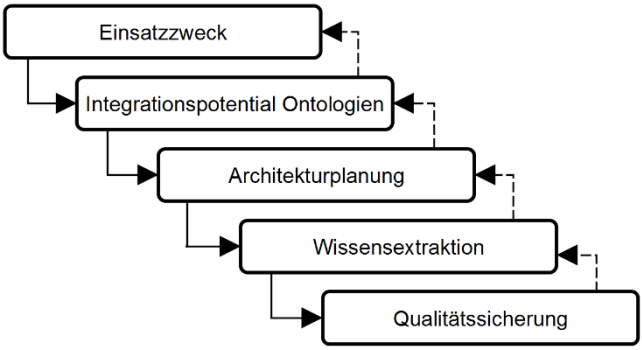


Abb. 1: Entwicklungsschritte bei der Entwicklung einer Domänenontologie.

2.2 Einsatzzweck

Im ersten Entwicklungsschritt werden Umfang und die als relevant erachteten Kernbereiche der Ontologie festgelegt. Für die Festlegung des Ontologieumfangs und der Modellierung von Anwendungsfällen wurde eine Liste mit 173 informellen Kompetenzfragen (Competency Questions) entworfen, deren Antworten später durch die Ontologie geliefert werden sollen. Für jede Frage werden außerdem die benötigten Konzepte festgehalten, die für ihre Beantwortung notwendig sind. Eine Frage aus dem Bereich des strategischen IM lautet beispielsweise: “Welche Geschäftsprozesse nutzen welche Anwendungssysteme?“ Das Vorgehen mit der Entwicklung von Kompetenzfragen bietet zwei entscheidende Vorteile. Einerseits lässt sich der abzudeckende Umfang relativ schnell auf bestimmte Kernbereiche eingrenzen. Auf der anderen Seite kann mit Hilfe der Fragen die Qualität der Ontologie (Schritt: Qualitätssicherung), durch korrekte Bereitstellung der gewünschten Informationen, sinnvoll getestet werden. Für eine Ontologie des Informationsmanagements im Krankenhaus wurden unter anderem die Bereiche „Strategisches IM“, „Taktisches IM“ und „Operatives IM“ bestimmt.

2.3 Integrationspotential vorhandener Ontologien

Ein großer Vorteil von Ontologien liegt in der Wiederverwendung bereits modellierten Wissens. Bei der Entwicklung einer Domänenontologie gestaltet sich dieser Punkt jedoch recht schwierig, da die Formen der Repräsentation von Domänenwissen unmittelbaren Einfluss auf die Anwendbarkeit bestimmter Methoden haben [BC88]. Die Wiederwendung und Integration auf rein konzeptueller Ebene ist allerdings durchaus möglich [GU97]. Für die Recherche nach geeigneten Architekturen auf konzeptueller

Ebene bieten sich verschiedene Plattformen (z.B. Linked Open Vocabularies[7], Ontolingua Ontology Library[8], DAML Ontology Library[9]) mit bereits existierenden, frei verfügbaren Ontologien verschiedener Themenbereiche. Für die Architektur der hier vorgestellten Ontologie finden sich Ansätze aus der „Core Enterprise Ontology" [BKM01] und insbesondere aus dem „3-Ebenen-Meta-Modell" (3LGM²) [WBW03] zur Beschreibung, Bewertung und Planung von Informationssystemen im Gesundheitswesen.

2.4 Architekturplanung

Im Vorfeld der eigentlichen Wissensextraktion empfiehlt es sich, bereits grundlegende Entscheidungen zur Architektur der geplanten Ontologie zu treffen. Bei diesem Schritt liegt der Fokus auf dem geplanten Einsatzzweck der Ontologie, der durch die Analyse der bereits entwickelten Kompetenzfragen unterstützt wird. Zu Beginn sollten geeignete Top-Level-Konzepte bestimmt werden, die die Begriffswelt der Domäne möglichst gut abdecken und nach welchen unterschieden werden soll. Dadurch gelingt gerade bei sehr großen Ontologien (>1000 Konzepte) eine gute Strukturierung der Begriffe und ermöglicht frühzeitig die Identifikation benötigter Basisrelationen zwischen den Top-Level-Konzepten. Für die Nutzung in zwei verwandten, dennoch verschiedenen Anwendungsszenarien (entscheidungsunterstützendes Werkzeug für den CIO, Lehre) sollten bereits jetzt alle erforderlichen Eigenschaften bestimmt werden, die während des Wissensextraktionsprozesses erfasst werden sollen. Bei diesem Schritt empfiehlt sich große Sorgfalt, da umfassende Architekturanpassungen während der Analyse- und Extraktionsphase zwar grundsätzlich möglich, jedoch unter Umständen mit großem Aufwand verbunden sind.

2.4.1 Die Top-Level-Ebene

Auf der Top-Level-Ebene unterscheiden wir für unsere Ontologie zunächst zwischen Aufgaben (functions) und Objekttypen (entities) [HAW05]. Bei Objekttypen wird zusätzlich zwischen aktiven Objekttypen (active entity type) und passiven Objekttypen (passive entity type) unterschieden, deren Verwendung im Folgenden näher erläutert wird [BKM01]. Das Konzept Aufgabe wird für unser Anwendungsgebiet weiter auf fünf krankenhausspezifische Top-Level-Aufgaben spezialisiert, welchen sich alle im Krankenhaus anzutreffenden Unternehmensaufgaben zuordnen lassen:

Aufgabe
- Unternehmensaufgabe
  - Krankenhausaufgabe
    - Aufgabe der Patientenversorgung
    - Aufgabe der Administration

[7] Linked Open Vocabularies: http://lov.okfn.org/dataset/lov/
[8] Ontolingua Ontology: http://www-ksl.stanford.edu/knowledge-sharing/ontologies/
[9] DAML Ontology Library: http://www.daml.org/ontologies/

- Aufgabe des Informationsmanagements
- Aufgabe der Managements
- Aufgabe der Ausbildung- und Forschung

Bei passiven Objekttypen handelt es sich um Informationsobjekte, die von Aufgaben für deren Erfüllung benötigt oder bearbeitet werden. Um später Aussagen über die von bestimmten Krankenhausaufgaben benötigten Informationsobjekte machen zu können, orientiert sich die Einteilung der passiven Objekttypen an den bereits bestimmten Aufgaben.

Passiver Objekttyp
- Unternehmensobjekttyp
  - Krankenhausobjekttyp
    - Objekttyp der Patientenversorgung
    - Objekttyp der Administration
    - Objekttyp des Informationsmanagements
    - Objekttyp der Managements
    - Objekttyp der Ausbildung und Forschung

Bei aktiven Objekttypen handelt es sich Rollen oder Personengruppen, die aktiv in Relation zu anderen passiven Objekttypen oder Aufgaben stehen. Es wird zwischen Rollen (role) und Abteilungen (department) unterschieden.

2.4.2 Passive und aktive Objekttypen

Für den Bereich „Informationsmanagement im Krankenhaus“ mit Blick auf die beiden geplanten Anwendungsgebiete wurden folgende Eigenschaften bestimmt, die für jeden Begriff (Passive und aktive Entitäten) in die Wissensbasis aufgenommen werden sollen:

- **Typ:** Jeder Begriff ist entweder vom Typ „Aufgabe“, „Objekttyp“ oder „Rolle“ - je nach Art seiner Zugehörigkeit. Spezifische Aufgaben des Informationsmanagements (Aufgabe IM), die sich von typischen Aufgaben innerhalb des Krankenhauses unterscheiden, werden als solche gekennzeichnet.
- **Domäne:** Für jeden extrahierten Begriff wird festgehalten, ob er für die Domäne Krankenhaus bzw. Gesundheitswesen spezifisch (gesundheitswesenspezifisch) ist.
- **Quelle:** Die extrahierten Aussagen nach dem RDF-Statement-Schema Subjekt-Prädikat-Objekt (z. B. „CIO“-„isResponsibleFor“-„Strategic HIS Planning“) sollen ihrem Ursprung zugeordnet werden können. Für jede Aussage werden daher sowohl Literaturquelle, als auch die entsprechende Seitenzahl notiert. Die Angabe der Literaturquelle bietet die Möglichkeit, die Ontologie kontinuierlich um weitere Literatur zu erweitern.

- **Definition:** Für den Einsatz der Ontologie in der Lehre sollte jeder Begriff über eine Definition verfügen. Durch die Anreicherung der Begriffe mit natürlicher Sprache wird die Verständlichkeit der Begriffszusammenhänge unterstützt.
- **Synonym:** Die Begriffsvielfalt im Bereich des IM wird durch die Zuordnung synonymer Bezeichner oder von Abkürzungen in Einklang gebracht. Bei der Suche in der Ontologie kann sowohl der Begriff als auch seine Synonyme als Suchbegriff verwendet werden.

2.4.3 Top-Level-Relationen

Nach Festlegung der Top-Level-Ebene enthält der nächste Schritt die Identifikation relevanter Top-Level-Relationen. Die in Abbildung 2 dargestellten Relationen wurden während des Extraktionsprozesses eingeführt und bei Bedarf kontinuierlich um spezifischere Relationen erweitert.

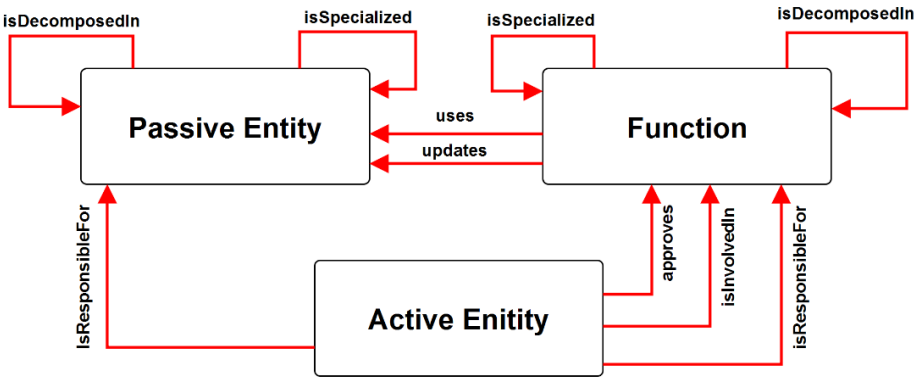


Abb. 2: Auszug aus der Spezifikation der Ontologiearchitektur. Mit Hilfe der drei Top-Level-Konzepte „Passive Entity“, „Active Entity“ und „Function“ lassen sich die benötigten Basisrelationen bestimmen.

Zugriff/Bearbeitung: A „uses“ B / A „updates“ B

Bei Konzept A handelt es sich um ein Konzept des Typs „Function“, bei Konzept B um ein Konzept des Typs „Passive Entity“. Die Relation „uses“ bedeutet, dass bei Durchführung von Aufgabe A Daten über den passiven Objekttyp B benötigt und genutzt werden. Die Relation „updates“ bedeutet, dass bei Durchführung von Aufgabe A Daten über den Objekttyp B erzeugt oder verändert werden.

Bsp.: Appointment Scheduling uses Case
Strategic Alignment updates Strategic Information Management Goal

Beteiligung/Verantwortlichkeit: A „isInvolvedIn“ B / A „isResponsibleFor“ B

Bei Konzept A handelt es sich um ein Konzept des Typs „Active Entity“, bei Konzept B entweder um ein Konzept des Typs „Passive Entity“ oder „Function“. Die Relation „isInvolvedIn“ drückt aus, dass A in B involviert ist (z.B. teilnimmt oder operativ mitarbeitet). Die Relation „isResponsibleFor“ drückt aus, dass A für B verantwortlich ist, i.d.R. sollte es sich bei Konzept A um eine Person bzw. eine Rolle einer Person handeln.

Bsp.: Consultant isInvolvedIn Long-Term HIS Planning
CIO isResponsibleFor Strategic Information Management Plan

Zustimmung: A „approves“ B

Bei Konzept A handelt es sich um ein Konzept des Typs „Active Entity“ (Person, Organisationseinheit), bei Konzept B um eine Aufgabe oder einen passiven Objekttyp. Die Relation drückt aus, dass der aktive Objekttyp A seine Zustimmung zu einer Aufgabe/passiven Objekttyp B gibt oder potenziell geben kann.

Bsp.: Hospital Management approves Strategic Information Management Plan

2.5 Wissensextraktion in drei Phasen

Die Extraktion der erforderlichen Informationen kann, wie in Abbildung 3 dargestellt, in drei Phasen durchgeführt werden, die sich teilweise parallelisieren lassen. In der Analyse- und Extraktionsphase wurden durch Zweierteams zunächst unabhängig voneinander kapitelweise Aussagen zu Aufgaben und Objekttypen des Informationsmanagements extrahiert und in Excel-Tabellen festgehalten. In der Konsolidierungsphase entstand im Zweierteam pro Kapitel ein abgestimmtes Extraktionsprotokoll. In der Validationsphase wurden die Einzelprotokolle durch ein Expertenteam auf Korrektheit überprüft und schließlich in ein Gesamt-Extraktionsprotokoll überführt. Dieser Schritt ist nochmals mit hohem Aufwand verbunden, da hier ein einheitliches Begriffssystem geschaffen werden muss und die Ontologiehierarchie aufgebaut wird. Mit Hilfe des selbstentwickelten Werkzeugs „Excel2OWL“ [Ta14] kann die einheitliche Endfassung in ein standardisiertes Ontologieformat (z.B. OWL) transformiert werden und lässt sich somit problemlos mit gängigen Ontologieeditoren analysieren und weiter modifizieren.

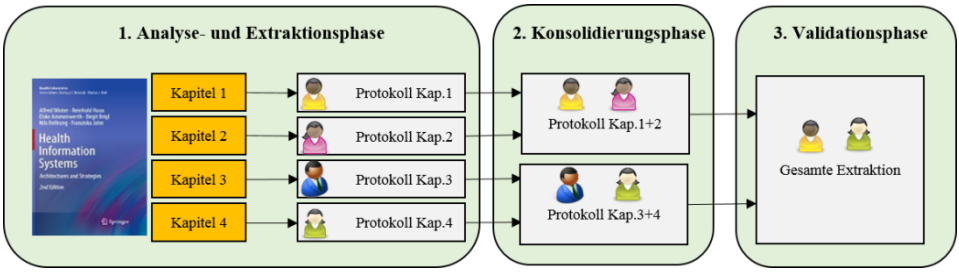


Abb. 3: Die Wissensextraktion wurde in drei Phasen eingeteilt. Der Analyse- und Extraktionsphase folgt unmittelbar eine Konsolidierung und Validation der Extraktionsergebnisse.

2.6 Qualitätssicherung

Die Erstellung großer Ontologien erfordert deren Wartung und Qualitätssicherung und muss auf mehreren Ebenen durchgeführt werden. Zum einen muss sichergestellt sein, dass die zuvor getroffenen Restriktionen und Vereinbarungen für Konzepte und Relationen beim Extraktionsprozess eingehalten wurden, zum anderen sollte sich das noch unformalisierte Wissen später auch formalisiert in der Ontologie befinden und abfragbar sein. Für beide Szenarien eignet sich die Überprüfung der Ontologie mit Hilfe der graph-basierten Abfragesprache für RDF-Formate SPAQRL[10]. Mit Hilfe eines SPARQL-Skripts kann z.B. die korrekte Verwendung der Relationen oder die einheitliche Zuordnung bestimmter Eigenschaften überprüft werden. Besonders geeignet sind die Anfragen bei sehr großen Ontologien mit vielen Relationen und Eigenschaften, da sehr schnell geprüft werden kann, wo Fehler aufgetreten sind oder Informationen vergessen wurden. Mit Hilfe der zu Beginn des Entwicklungsprozesses erstellten Kompetenzfragen lässt sich die inhaltliche Qualität der Ontologie überprüfen. Die Kompetenzfragen werden hierfür in entsprechende SPARQL-Anfragen transferiert und die Ontologie abgefragt. In Abbildung 4 ist das Ergebnis einer Abfrage im Rahmen der Qualitätssicherung dargestellt. Ein unabhängiger Domänenexperte bewertet im Anschluss die Qualität der Antworten.

```
[Exel2OWL] QS-Kriterium3: Nur Aufgaben dürfen die Relation UPDATES verwenden
[Exel2OWL] Zeige alle Subjekte die dieses QS-Kriterium nicht erfüllen:
-----------------------------------------------------------------------------------------------------------
| Subject                                          | Subjecttype    | Object                                |
===========================================================================================================
| "Administrative_Discharge_and_Billing"           | "Entity Type" | "Bill"                                 |
| "Administrative_Discharge_and_Billing"           | "Entity Type" | "Case"                                 |
| "Administrative_Discharge_and_Billing"           | "Entity Type" | "DRG"                                  |
| "Administrative_Discharge_and_Billing"           | "Entity Type" | "Notification"                         |
| "Approval_to_Strategic_Information_Management_Plan" | "Entity Type" | "Strategic_Information_Management_Plan" |
| "DRG_Calculation"                                | "Entity Type" | "DRG"                                  |
| "Order_Entry"                                    | "Entity Type" | "Diagnostic_Procedure"                 |
| "Order_Entry"                                    | "Entity Type" | "Therapeutic_Procedure"                |
-----------------------------------------------------------------------------------------------------------
```

Abb. 4: Ergebnis einer SPARQL-Anfrage zur Qualitätssicherung der Ontologie. Das Ergebnis zeigt alle falsch abgebildeten Aussagen der Relation „updates“.

[10] SPARQL Protocol and RDF Query Language

3 Ergebnisse

Die vorgeschlagene Methode zur Extraktion von fachspezifischem Wissen und formalen Repräsentation in einer Ontologie wurde am Beispiel des Lehrbuchs „Health Information System: Architectures and Strategies“ vorgestellt. Insgesamt konnten 1121 Konzepte und über 2255 Relationen extrahiert und abgebildet werden, die die komplexen Zusammenhänge des Informationsmanagements in Krankenhäusern beschreiben. Darüber hinaus können für einen Großteil der Konzepte zusätzliche Informationen wie Definitionen, Synonyme, alternative Bezeichnungen und Verweise auf die Literaturquelle abgerufen werden. Für die Repräsentation der Ergebnisse wurde der weit verbreitete Open-Source Editor Protégé[11] zur Modellierung und Visualisierung von Ontologien eingesetzt. Die Ontologie des Informationsmanagements im Krankenhaus lässt sich nun auf verschiedene Arten nutzen. Durch die Bereitstellung zahlreicher Protégé-Plugins (Ontograf, OntoViz) kann die modellierte Wissensbasis frei skalierbar dargestellt werden. Die graphische Visualisierung der Ontologie bietet eine hervorragende Möglichkeit zwischen den modellierten Begriffen zu navigieren und sie nach eigenem Ermessen anzuordnen, Sichten zu skalieren oder den Fokus auf die Verwendung bestimmter Relationen zu lenken. Für die verschiedenen Kernbereiche lassen sich so ganze Begriffslandkarten individuell gestalten und abspeichern.

Die Prüfungsvorbereitung für Studenten oder auch die Prüfungsdurchführung stellt einen weiteren Anwendungsfall für die Verwendung der Ontologie dar. Studenten im Prüfungsmodel „Management von Informationssystem im Gesundheitswesen“ sehen sich beispielsweise mit folgender Prüfungsfrage konfrontiert: Wie ist ein strategischer Rahmenplan aufgebaut und wer ist für ihn verantwortlich? Während die Antworten auf diese Frage verstreut über einige Kapitel im Lehrbuch zu finden sind, bietet die Ontologie effektivere Zugriffsmechanismen. Die Antwort auf die Frage lässt sich einerseits über die Suchfunktion von Protégé ermitteln oder über eine entsprechend generierte SPQARL-Anfrage an die Ontologie. In Abbildung 4 ist das Ergebnis der Anfrage zu sehen mit den entsprechenden Antworten zu sehen. Der strategische Rahmenplan besteht offensichtlich aus der Beschreibung des aktuellen und geplanten Zustands des Krankenhausinformationssystems (KIS), einer Analyse und Bewertung des aktuellen KIS, einem strategischen Krankenhausziel, dem strategischen IM-Ziel und einem Migrationspfad. Die Verantwortung liegt beim CIO.

[11] http://protege.stanford.edu/



Abb. 5: Ergebnis der Anfrage nach den Bestandteilen und der Verantwortlichkeit des „Strategic Information Management Plan“ an die Ontologie. Die gelb-gestrichelten Linien zeigen die Komponenten des Plans über die Teil-von Beziehung. Die Verantwortlichkeitsrelation wird durch die braun-gestrichelte Linie dargestellt.

Um detailliertere Informationen (Definition, Synonyme…) zu einzelnen Konzepten zu erhalten, bietet Protégé die Möglichkeit, die bei der Wissensextraktion erhobenen Informationen durch Anklicken des Konzepts zu erhalten. Dadurch lassen sich schnell und übersichtlich alle modellierten Informationen zu jedem Konzept abrufen. Eine weiterführende Prüfungsfrage könnte nun auf die Verantwortlichkeiten eines CIOs abzielen: Für welche Aufgaben ist der CIO in einem Krankenhaus verantwortlich?

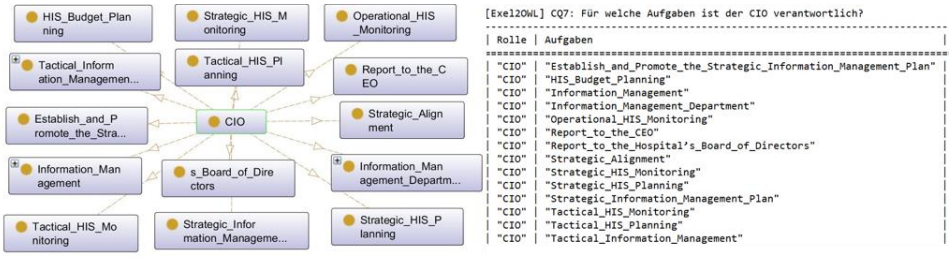


Abb. 6: Ergebnis der Anfrage nach den Verantwortlichkeiten eines CIO. Rechts: Visuelle Darstellung der Lösung mit Ontograf. Links: Darstellung der Ergebnisse unter Verwendung von SPARQL-Anfragen.

4 Diskussion und Ausblick

Die Ontologie Informationsmanagements im Krankenhaus soll sowohl als Basis für die akademische Lehre als auch bei der Entwicklung entscheidungsunterstützender Werkzeuge dienen. Für die geplante Verwendung in zwei Bereichen müssen deren Anforderungen an die Ontologie regelmäßig geprüft werden und Architekturentscheidungen gegebenenfalls angepasst werden. Der erste Schritt zur Entwicklung einer Ontologie des Informationsmanagements im Krankenhaus ist durch die formale Abbildung des Lehrbuchs „Health Information Systems – Architectures and

Strategies“ gemacht. Die Ergebnisse sind vielversprechend und bilden unser Verständnis für die Domäne des Informationsmanagements im Krankenhaus in Form einer formalen Ontologie ab. Für die Vernetzung mit weiteren theoretischen Ansätzen und Frameworks wie COBIT und ITIL müssen nun weitere Wissensquellen an die hier vorgestellte Ontologie angebunden werden. Die folgenden Entwicklungsschritte sehen nun die Erweiterung der Ontologie um Themengebiete wie „IT-Projektmanagement“, „Vertragsmanagement“ und „IT-Servicemanagement“ vor. Denn erst durch die Anbindung weiterer Wissensquellen kann die Vernetzung ihrer Begriffswelten realisiert werden.

Danksagung

Die vorliegende Arbeit ist Teil des durch die Deutsche Forschungsgemeinschaft (DFG) geförderten Forschungsprojekts „Semantisches Netz des Informationsmanagements im Krankenhaus (SNIK)“, Förderungsnummern: 1605/7-1 (Leipzig), 1387/8-1 (Heidelberg).

Literaturverzeichnis

[MSM08] Messerschmidt, M.; Schülein, P.; Murnleitner, M.: Der Wertbeitrag der IT zum Unternehmenserfolg: Manage IT as a business. PricewaterhouseCoopers AG, Stuttgart, S.6, 2008.

[JTP11] Jaana, M.; Tamim, H.; Paré G.; Teitelbaum M.: Key IT-Management issues in hospitals: Results of a Delphi study in Canada. International Journal of Medical Informatics 2011 Dec;80(12):828-40, 2011.

[Kr09] Krcmar, H.: Informationsmanagement. 5. Auflage, Springer, Berlin, 2009.

[Wi11] Winter A.; Haux R.; Ammenwerth E.; Brigl B.; Hellrung N.; and Jahn F.: Health Information Systems: Architectures and Strategies. 2nd ed., Springer, Heidelberg, 2011.

[GFC04] Gómez-Pérez, A.; Fernández-López, M.; Corcho, O.: Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Springer-Verlag, London, Berlin, Heidelberg, 2004.

[NM01] Noy, N. F.; McGuinness D. L.: Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, 2001.

[FGJ97] Fernandez, M.; Gomez-Perez, A.; Juristo N.: Methontology: From Ontological Art Towards Ontological Engineering. Technical Report AAAI Report No.: SS-97-06, 1997.

[Ja14] Jahn, F.; Schaaf, M.; Paech B.; Winter A.: Ein semantisches Netz des Informationsmanagements im Krankenhaus. GI-Jahrestagung 2014, Stuttgart, 1491-1498.

[BC88] Bylander T.; Chandrasekaran B.: Generic tasks for knowledge-based reasoning: the "right" level of abstraction for knowledge acquisition. International Journal of Man-Machine Studies - Knowledge acquisition for knowledge-based systems. Part 2, Volume 26 Issue 2, 231 – 243, 2008.

[Gu97] Guarino, N.: Understanding, Building, and Using Ontologies: A Commentary to "Using Explicit Ontologies in KBS Development", by van Heijst, Schreiber, and Wielinga. International Journal of Human and Computer Studies (46): 293-310, 1997.

[BKM01] Bertolazzi, P.; Krusich, C.; Missikoff, M.: An approach oft he definition of a core enterprise ontology: CEO. Paper presented at the International Workshop on Open Enterprise Solutions: Systems, Experiences and Organizations (OES-SEO 2001), Rome, 14.-15.09.2001.

[WBW03] Winter, A.; Brigl, B.; Wendt, T.: Modeling Hospital Information Systems (Part 1): The Revised Three-layer Graph-based Meta Model 3LGM². Methods of Information in Medicine, 42 (5), 544-551, 2003.

[HAW05] Huebner-Bolder, G.;Ammenwerth, E.; Winter, A.; Brigl B.: Specification of a Reference Model for the Domain Layer of Hospital Information Systems In: Engelbrecht R, Geissbuhler A, Lovis C, Mihalas G (Hrsg.): Connecting Medical Informatics and Bio-Informatics. Proceedings of Medical Informatics Europe (MIE 2005), Geneva, Aug 08 - Sep 01 2005. Studies in Health Technology and Informatics, Volume 116. Amsterdam: IOS Press, 497-502, 2005.

[Ta14] Tahar K.; Jahn F.; Schaaf M.; Kücherer C.; Paech B.; Winter A.: Eine Ontologie für die Unterstützung der Lehre und des Informationsmanagements im Gesundheitswesen. 13th Leipzig Research Festival for Life Sciences, 45p, Leipzig, 2014.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Tiefenanalyse zum Performance Management in Schweizer Krankenhäusern auf Basis dreier Fallstudien

Konrad Walser[1], André Meister[2], Etienne Huber[3]

Abstract: Ausgehend von einer quantitativen Untersuchung zum Performance Management in Schweizer Krankenhäusern, einschließlich der Teilthemen Planung, Controlling, Reporting sowie IT-Unterstützung ausgehend von der DRG-Einführung, interessierte die Frage, welche Reife die Krankenhäuser in den erwähnten Themenbereichen haben (Stand 2012 September) und welche geplante Reife für den September 2014 vorgesehen war. Die Ergebnisse überraschten auf der einen Seite, zeigten aber auf der anderen Seite auch auf, dass die Krankenhäuser in der Schweiz einen guten Stand und mehrheitlich klare Vorstellungen zur Weiterentwicklung in den genannten Themen haben. Für den vorliegenden Beitrag wurde ausgehend davon und mittels qualitativer Forschung mittels dreier Fallstudien zu unterschiedlich großen Krankenhäusern in der Deutschschweiz die Ausprägung des Performance Managements in die Tiefe gehend untersucht. Aus der früheren quantitativen Studie und deren Ergebnisse konnten diverse Hypothesen abgeleitet werden, die ergänzt wurden durch ein Integrationsmodell für Krankenhäuser, anhand dessen Integrationsrichtungen und Schwerpunkte im Zeitverlauf eruiert werden können. Die Resultate aus dem Vergleich der drei Fallstudien erbrachte folgende Ergebnisse. Im einen großen Krankenhaus erfolgt eine Konzentration auf organisatorische Maßnahmen vor technischer Implementierung. Der Aufbau von technischem Know-how um Performance-Management über Data Warehousing- und Business-Intelligence-Technologie zu unterstützen ist nachgelagert aber im Gange. Das Business-IT-Alignment ist in ebendiesem Krankenhaus nicht perfekt; dafür verstehen sich Medizin und Administration gut und es besteht keinerlei Dissens über die Ausrichtung des Hauses zwischen medizinischer und administrativ-betriebswirtschaftlicher Führung. Das Haus ist sehr gut positioniert was das Performance Management und dessen Wirkung betrifft und hat hohe betriebswirtschaftliche Ansprüche. Im zweiten kleinen Haus hat die DRG-Einführung zu organisatorisch-technischen Veränderungen geführt. Sehr deutlich wird an dieser Fallstudie die intensive Suche nach der Beeinflussung der Fälle (DRG‘s) und der Professionalisierung der Abrechnung sowie der Optimierung des Managements der DRG’s und der dahinterliegenden Cases. Möglicherweise herrscht bezüglich des Performance Managements der Kleinheit des Hauses halber ein Economies-of-Scale-Problem vor. Der Wettbewerb schlägt im Beispiel dieser Fallstudie voll zu. Das größte untersuchte Haus kämpft mit der Spezialisierung, der dadurch entstehenden zusätzlichen Kosten für Cases, welche die Base-Rates meist überschreiten. Zudem ist dem Haus die Größe hinderlich. Diese lässt die Dichotomie innerhalb der Geschäftsleitung (partiell) sowie der Direktionen/Klinika bezüglich Medizinischer und betriebswirtschaftlich Leitung voll zutage treten. Es ist aber das einzige der untersuchten Krankenhäuser, in dem der CIO zur damaligen Zeit in der GL vertreten war.

Keywords: Performance Management, DRG’s, Neue Krankenhausfinanzierung Schweiz, ICT, Data Warehousing, Reporting, Krankenhäuser, Informationsmanagement im Krankenhaus.

[1] Berner Fachhochschule, E-Government-Institut, Brückenstrasse 73, CH-3005 Bern, konrad.walser@bfh.ch
[2] NOVO Business Consultants, Gutenbergstrasse 50, CH-3011 Bern, andre.meister@novo-bc.ch
[3] Eidgenössisches Departement für Verteidigung, Bevölkerungsschutz und Sport, Bamako/Mali, etienne.huber@gmail.com

1 Einleitung

1.1 Problemstellung

Die Krankenhäuser in der Schweiz stehen ausgehend von der DRG-Einführung unter hohem Handlungsdruck ([Af13], [NZZ11], [Og12]). Es herrscht zunehmender Wettbewerb. Die Frage der organisatorischen Positionierung und der Veränderung im Hinblick auf den Wettbewerbsdruck beherrscht alle Häuser in ähnlichem Maße. Ausgehend von Analysen zum Performance Management (PM) in Schweizer Krankenhäusern ([WMH13], [WHMoJ]) kann folgende Kausalkette propagiert werden: Die DRG-Einführung führt neben der neuen Krankenhausfinanzierung als Ursache zur Veränderung hin zu mehr PM und organisatorischen Auswirkungen. Dies hat Auswirkungen auf das System der „Performance Steuerung“ bis hin zur Unterstützung des PMs durch IT-Instrumente. Dabei ist der Zusammenhang zwischen der Veränderung einzelner Instrumente wie Budgetierung, Planung, Controlling sowie IT-Unterstützung besonders zu beachten. Die folgenden Fragen können ausgehend davon gestellt werden: Wie wirkt sich dieser Handlungsdruck in Richtung PM organisatorisch und IT-spezifisch aus (Wettbewerb, Beweglichkeit der Organisation im Hinblick auf Wettbewerb, interne organisatorische Veränderungen, Informationsintegration zum PM, etc.). Welche Auswirkungen hat die Einführung des PMs auf die IT und deren Integration? Die bereits erwähnte quantitative Untersuchung ([WMH13], [WHMoJ]) führte zu den folgenden Resultaten: Mit der Einführung von Wettbewerbsmechanismen (DRG- und Einführung eines neuen Krankenhausfinanzierungssystems) im neuen Krankenversicherungsgesetz [BAG13] wird die betriebswirtschaftliche Ausrichtung der Krankenhausführung zum Schlüsselfaktor für ein Bestehen am Markt. Die deskriptive Auswertung der Daten der quantitativen Studie zeigte, dass die meisten Schweizer Krankenhäuser auf dem richtigen Weg sind, jedoch die strategische Wichtigkeit eines wirksamen PMs vielerorts noch unterschätzt wird. Der IT kommt als Plattform des PMs eine zentrale Rolle zu. Die Krankenhausplanung stellt den eigentlichen Treiber zum Auf- und Ausbau von leistungsfähigen PM-Systemen dar. Erst eine gut entwickelte IT und klare Planwerte lassen ein inhaltlich umfangreiches PM zu. Daneben darf nicht vergessen werden, dass ein funktionierendes PM auch eine organisatorische Komponente beinhaltet. Erst wenn AKV im Hinblick auf die Krankenhaussteuerung bezüglich Betriebswirtschaft, Medizin und Pflege aufeinander abgestimmt sind und harmonieren, können Krankenhäuser leistungsorientiert geführt werden. Nicht weiter betrachtet wurden in der eigenen quantitativen Untersuchung die folgenden Determinanten für ein erfolgreiches PM: Kultur, Marktdruck, Dichotomie Expertenorganisation (Medizin/Pflege) versus Administration/betriebswirtschaftliche Führung des Krankenhauses sowie das Business-IT-Alignment.

1.2 Zielsetzung

Die folgenden Zielsetzungen werden mit der vorliegenden qualitativen Untersuchung

verfolgt: Ableitung eines Hypothesenmodells aus einem Wirkungsmodell, das ausgehend von den Nachbereitungen zur erwähnten quantitativen Untersuchung ([WHM13], [WHMoJ]) entstand; Umsetzung der Hypothesen in einen Interviewleitfaden; Prüfung der Hypothesen anhand dreier Fallstudien; Konkretisierung von Schlussfolgerungen aus den Fallstudien. Im Besonderen wurde den Aspekten, die in der quantitativen Studie nicht untersucht wurden (vgl. Absatz 1.1), in der qualitativen Studie besondere Bedeutung zugemessen.

1.3 Methodisches Vorgehen

Ausgehend von der Problemstellung und den Zielen interessiert ein tieferer Einblick in Krankenhausinterna zur Eruierung der Zusammenhänge, die zu einem mehr oder weniger erfolgreichen PM führen. Dafür sind Fallstudien klar die bessere Untersuchungsmethode als quantitative Untersuchungen, was zur vorliegenden Untersuchung anhand dreier Fallstudien führte. Ausgehend von den Resultaten der früheren quantitativen Untersuchung erfolgte auch eine Konkretisierung von Hypothesen. Diese konnten wiederum in einem Ursache-Wirkungs-Modell des Krankenhauses verdichtet werden. Daraus erfolgte die Ableitung von Fragen, die im Rahmen der Fallstudien adressiert wurden (Interviewleitfaden im Anhang). Zu den methodischen Aspekten der Fallstudien wurde [Yi09] beigezogen. Überdies wurde zur Steigerung von Validität und Reliabilität durch Triangulation für die vorliegende Studie das folgende Vorgehen gewählt [Yi09]. Die Evidenz ist ausgehend von der quantitativen Umfrage gegeben. Das Thema erscheint relevant. Es wurden explizit drei Fallstudien aus den auf die quantitative Umfrage antwortenden Akut-Krankenhäusern ausgewählt. In einem Fall wurde ein Geschäftsbericht zur Validierung der Aussagen aus dem Interview beigezogen. Überdies wurde (aus Triangulationsgründen) ein (nicht dokumentiertes) Interview mit einem ehemaligen Pharmavertreter für Krankenhäuser gemacht, ausgehend von dem insbesondere Bestätigungen bezüglich der Typologie der Fallstudien von Akut-Krankenhäusern und den Schlussfolgerungen erreicht wurden, d.h. eine Cross-Validierung der Auswertungen. Überdies konnte die Triangulation durch Verifizierungen der Fallstudien durch die Interviewten erweitert werden (ein- bis mehrmaliges Gegenlesen). Die Niederschrift der Interview-Antworten erfolgte möglichst nahe der schriftlichen Interviewdokumentation entlang (ohne Tonbandaufnahmen und entsprechende Transkriptionen). Eine Inhaltsanalyse erschien ausgehend von den Interviewfragen nicht zielführend. Es ging in den Interviews nicht um eine Verifizierung von Einstellungen beispielsweise auf Basis einer Skala, sondern vielmehr um die konkrete Sachverhaltseruierung strategischer, organisatorischer und technischer Art hinsichtlich des „PM-Systems Krankenhaus“. Die Aufnahme der Fallstudien erfolgte durch einen der drei Autoren. Eine weitere Triangulationsmöglichkeit ergab sich danach dadurch, dass die Coautoren die resultierenden Texte verifizierten und hinsichtlich Konsistenz bezüglich Schlussfolgerungen prüften. Ausgehend von den quantitativen Untersuchungen und ausgehend von [Yi09] resultierte ein unter den Autoren abgesprochenes Set an drei Fallstudien (kleines, mittelgroßes, sehr großes Akutkrankenhaus). Akut-Krankenhäuser wurden gewählt, weil diese am direktesten von der Einführung von DRG’s und Neuer Krankenhausfinanzie-

rung betroffen sind. Man hätte auch drei gleich große Häuser wählen können, welche sich vom Behandlungsangebot her unterschieden hätten oder auch nicht. Allerdings, dies ist als Limitation anzumerken, kann es sein, dass jedes der gewählten Krankenhäuser für seine Größenkategorie untypisch ist, was möglicherweise einen Einfluss auf die Schlussfolgerungen haben kann. Eine Anonymisierung erfolgte, um mit den Interviewten eine möglichst offene Interviewsituation zu erlangen und damit auch möglichst tiefe Einblicke in die PM-Problematik der entsprechenden Akutkrankenhäuser zu erhalten. Aus Wettbewerbssicht hätten Namensnennungen negative Konsequenzen bezüglich Auskunftsbereitschaft haben können. Die Grundüberlegung für die Auswahl unterschiedlich ausgeprägter Größen von Akut-Krankenhäusern lautete bezüglich der vorliegenden Untersuchung: Untersuchung des Größeneinflusses auf unterschiedliche Ausprägungen des PMs ausgehend von der DRG-Einführung; Untersuchung des Größeneinflusses auf die IT-Integrationsfähigkeit und -Maturität ausgehend von der DRG-Einführung; Untersuchung des Größeneinflusses auf die Fähigkeit, sich organisatorisch den neuen Steuerungsanforderungen anzupassen und die Steuerungsfähigkeit im Krankenhaus ausgehend von der DRG-Einführung in den Griff zu bekommen sowie Leichtigkeit der Ausprägung der organisatorischen Anpassung.

1.4 Ableitung der für die Fallstudienprüfung relevanten Hypothesen

Ausgehend von der Problemstellung und von der Beschreibung und Interpretation der quantitativen Umfrageresultate ([WMH13], [HWMoJ]) werden im Folgenden Hypothesen formuliert und deren Entstehung erläutert. Die Hypothesen dienen als Ausgangspunkte für die Analyse der drei Fallstudien. Damit werden weitergehende und vertiefende Analysen ermöglicht, welche über das für die quantitative Untersuchung genutzte Reifegradmodell im Sinne einer holistischeren PM-Sicht hinausgehen. Das neue Steuerungsparadigma ausgehend von der DRG-Einführung in den deutschschweizerischen Krankenhäusern erforderte organisatorische Veränderungen, die abhängig u.a. von der Ausrichtung und der Größe des Krankenhauses unterschiedlich ausfallen (so die Annahme). Folgende mögliche Fragen resultieren daraus: Welche Bedeutung erhalten betriebswirtschaftliche Steuerungsparadigmen in Krankenhäusern ausgehend von der DRG-Einführung? Wie geht das Krankenhaus aus organisatorischer Sicht damit um, etwa bei der Zusammensetzung der Geschäftsleitung, der Führungsstruktur und der Stärkung betriebswirtschaftlich-administrativer Stellen oder der Unter- oder Überordnung von Betriebswirtschaft, Medizin, Pflege und IT ([DC99], [CCT11])? Dies führt zu Hypothese 1: Wenn in Krankenhäusern ein PM-System aufgebaut werden soll, dann sind dafür primär organisatorische Änderungen und weniger technische Entwicklungen (Implementierungen neuer Informationssysteme) erforderlich. Organisatorische Änderungen sind langwierig, da Verhaltensweisen, Prozesse und Schnittstellen verändert werden müssen [Ch84]. Dafür ist ein strukturiertes Change Management nötig, dem technologische Änderungen im Idealfall nachfolgen. Von zentraler Bedeutung dabei ist, dass einerseits die finanziellen Ressourcen für technische Veränderungen vorhanden sind, dass andererseits ausgehend von der organisatorischen Veränderung die Anforderungen an die zu adaptierende IT präzise aufgenommen werden. Ausgehend davon macht die

Hypothese 2 explizit, dass eine temporale Abhängigkeit zwischen organisatorischer Veränderung und technischer Adaption des Informationssystems Krankenhaus und entsprechenden Informationstechnologien besteht: Wenn für eine neue Planungs- und Steuerungsphilosophie im Krankenhaus organisatorische Änderungen erforderlich sind, dann dauert die dafür erforderliche Organisationsentwicklung länger als die Anpassungen entsprechender Informationssysteme. Eine im Hinblick auf ein integriertes PM erforderliche Veränderung von Informationssystemarchitekturen hat strukturiert, d.h. auf Basis einer Unternehmensarchitektur zu erfolgen und dauert länger als die Integration (oder Einführung einzelner und allenfalls wenig integrierter) Informationssysteme für das PM. Von entscheidender Bedeutung bei der Neuausrichtung von Krankenhäusern ist das Bewusstsein für die Bedeutung des Business-IT-Alignments. Ein entsprechendes Business-IT-Alignment-Bewusstsein ist eher bei reiferen Organisationen vorhanden. Damit ist davon auszugehen, dass, je ausdifferenzierter eine IT (tendenziell eher bei größeren Krankenhäusern) ist, ein umso reiferes Business-IT-Alignment vorhanden ist. Dies führt dazu, dass die zusammenhängende Betrachtung von Organisations- und IT-Entwicklung verstanden und entsprechend gehandelt wird. Dies führt zur Hypothese 3: Wenn in Krankenhäusern entsprechende Organisationsänderungen erforderlich sind, dann ist sich das Krankenhaus darüber nicht bewusst und implementiert/adaptiert der Einfachheit halber eher IT als die Organisation. Die oben gemachten einleitenden Ausführungen zur Hypothese 3 führen zu einer weiteren Hypothese 4: Wenn die Maturität der IT in Schweizer Krankenhäusern relativ gering ist, dann werden eher einfache und wenig automatisierte Lösungen statt State-of-the-Art-Technologien für die Analyse von steuerungsrelevanten Informationen eingesetzt. Dies verunmöglicht adressatengerechte Auswertungen und Informationsbereitstellungen auf Basis von Informationssystemen. Der Primat der Steuerung des Krankenhauses (und dessen IT) ist wesentlich abhängig von der Zusammensetzung und Gewichtung der verschiedenen Fachbereiche in einem Krankenhaus in Relation zueinander über die strategische, taktische und operative Steuerungsebene. Dies führt zur Hypothese 5: Wenn die organisatorische Veränderung der Krankenhausorganisation von einer Expertenorganisation hin zu einer betriebswirtschaftlich geführten Organisation führen soll, dann müssen entweder die Experten zu stärkerem betriebswirtschaftlichem Denken angehalten werden oder aber die betriebswirtschaftlich orientierten Stellen im Krankenhaus müssen bezüglich Führungsfunktionen im Krankenhaus aufgewertet werden. Die Hypothese 6 postuliert einen Zusammenhang zwischen einer mangelnden betriebswirtschaftlichen Reife und einem mangelnden Verständnis für die Zusammenhänge zwischen strukturierter Entwicklung von IT und Organisation. Dies kann zur Illusion führen, dass IT-Änderungen und architektonische Änderungen alleine zur Neuorientierung beitragen können: Wenn die architektonische IT-Adaption im Hinblick auf die organisatorische Krankenhaussteuerungsperspektive im Vordergrund steht, dann stellt sich im Rahmen der Frage nach künftigen Veränderungen – aus einem Nichtbewusstsein bezüglich der Gewichtigkeit und der erforderlichen Dauer organisatorischer Änderungen – eine Selbstüberschätzung bezüglich technischer Veränderungsmöglichkeiten bei den Krankenhäusern ein. Trifft die Hypothese 6 zu, kann im Wesentlichen auch von einem mangelhaften gegenseitigen Verständnis bezüglich Möglichkeiten/Grenzen die durch IT für den

Krankenhausalltag über die Zeit resultieren (Mangelhaftes Business-IT-Alignment). Die Hypothese 7 thematisiert ähnlich wie die Hypothese 1 die Tatsache, dass eine stärker steuernde Krankenhausorganisation das neue DRG-Steuerungsparadigma unterschiedlich implementieren kann. Dies führt zur Hypothese 7: Wenn die Krankenhausorganisation zu einer betriebswirtschaftlich steuernden Organisation werden soll, dann muss die Maturität bezüglich Steuerung/Planung aller Krankenhausbereiche verbessert werden und nicht nur medizinisch oder betriebswirtschaftlich. Die Hypothesen sind in einem Ursache-Wirkungs-Modell wie folgt positionierbar (Abbildung 1).

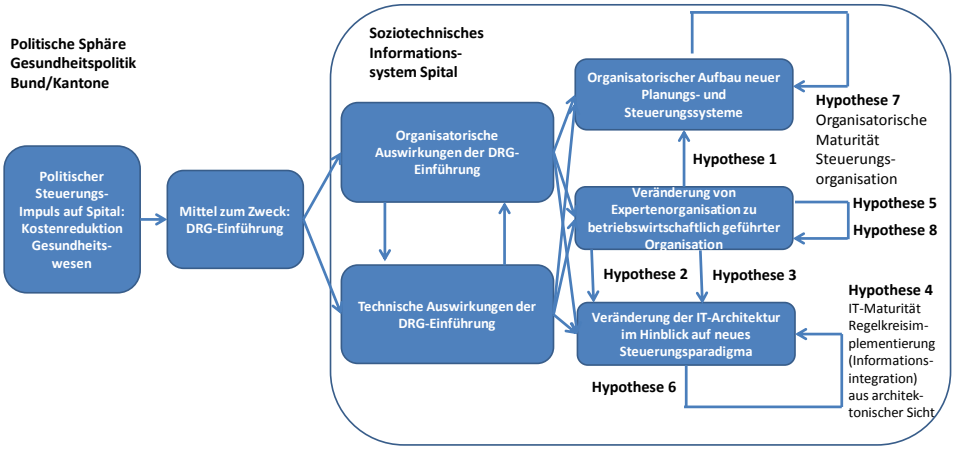


Abbildung 1: Positionierung der Hypothesen im sozio-technischen Informationssystem Krankenhaus.

Dabei wird vereinfachend davon ausgegangen, dass der (externe) politische Druck und Steuerungsimpuls zur verstärkten Kontrolle der Gesundheitskosten in der stationären Behandlung zur Einführung von DRG's und damit zur Änderung des Abrechnungssystems führt. Dies führt im Krankenhaus zu organisatorischen und technischen Anpassungen, welche die Häuser autonom vornehmen. Im Wesentlichen bedeutet Letzteres, dass verstärkt Datenintegrationen von operativen Systemen (KIS-, ERP-, Systeme, etc.) in dispositive Systeme (Data Warehouses, Business-Intelligence) erfolgen ([Me09], [MM09]), damit benutzergerechte Informationsbereitstellungen resultieren. Controller müssen überdies zu „Ermöglichern" für eine rollenspezifische Auswertung von Daten mutieren. Ausgehend davon sind einerseits organisational neue Planungs- und Steuerungssysteme und -prozesse einzuführen. Dies hat zur Folge, dass sich die Krankenhausleitung über die Geschäftsleitungszusammensetzung Gedanken machen muss. Überdies ist das Business-IT-Alignment zu fördern und zu entwickeln. Dies ermöglicht die Entwicklung der IT-Landschaft in Abhängigkeit von einem Entwicklungspfad der Geschäftsseite in Richtung einer betriebswirtschaftlich steuernden Organisation. Dabei werden Investitionen immer in Abhängigkeit von der Reife der Geschäftsseite bezüglich Planung, Budgetierung und PM getätigt.

2 Fallstudien

Die Interviews fanden mit je einer Person dreier Akut-Krankenhäuser der Schweiz statt. Ein tabellarischer Vergleich mit Fakten zu den Häusern lautet wie folgt.

	Krankenhaus A	**Krankenhaus B**	**Krankenhaus C**
Größe	Klein	Groß	Sehr groß
Örtlichkeit	Land (Agglomerations-Gebiet)	Mittelgroße Stadt	Große Stadt
Wettbewerb	Sehr intensiv	Eher weniger intensiv	In Spezialisierungs-Bereichen sehr intensiv
Fallarten	Einfachere Fälle	Komplexere und weniger komplexe Fälle	Vorwiegend komplexe Fälle
Größe der IT	Klein	Mittelgroß	Sehr groß
Interview-Daten	April/Mai 2013	April/Mai 2013	April/Mai 2013
Interviewpartner	Leiter Finanzen	Leiter Krankenhaus-Entwicklung	Leiter Finanzen
Kategorie: Anzahl stationärer Fälle im Jahr 2011	Zwischen 6000 – 8999	Zwischen 9000 und 30000	Größer 30000

Abbildung 2: Eingangsvergleich zu den drei Fallstudien.

2.1 Fallstudie A – Klein-Krankenhaus

Der Interviewte hatte zuletzt 2.8 Stellen an Mitarbeiterkapazität im Finanzbereich. Zu den Finanzen gehörten alle Zusatzbereiche, so etwa auch die Patientenadministration. Insgesamt hat der Bereich 29 Mitarbeiter. Die neue Pflegedienstleitung denkt und führt stark betriebswirtschaftlich. Dies steht zum Teil im Gegensatz zur Belegschaft. Die Ärzteschaft denkt nicht sehr betriebswirtschaftlich. Einer von sechs Chefärzten denkt auch stark betriebswirtschaftlich. Einflussfaktoren für die Veränderungen im Krankenhaus sind – neben der DRG-Einführung – der Wettbewerb. Neu sind auch Privatkrankenhäuser auf der Krankenhausliste. Veränderungen organisatorisch-technisch in den Krankenhäusern haben entsprechend stark zugenommen. Im Einzugsgebiet ist der Wettbewerb insbesondere mit privaten Häusern sehr intensiv. Die wesentliche Veränderung mit der DRG-Einführung war die Umstellung von Teil- auf Vollkostenrechnung. Das Krankenhaus muss nun, um seine Investitionen selber tätigen zu können, Gewinn machen. Vor der neuen Spitalfinanzierung investierten die beteiligten Gemeinden. Die kantonale Planungs- und Regelungsdichte nahm damit stark zu. Die Krankenhausplanung schreibt vor, welche Leistungen in der Grundversorgung angeboten werden. Krankenhäuser dürfen auch Spezialitäten anbieten, die bei der kantonalen Gesundheitsdirektion separat zu beantragen sind. Es werden Vorgaben bezüglich DRG's sowie u.a. auch bezüglich Codierung gemacht. Der Einfluss der DRG-Einführung auf die Organisation war groß. Die große Frage lautete für das Krankenhaus: Wie sind die Pa-

tienten/DRG's steuerbar? Aus Wettbewerbssicht liegt das Krankenhaus im Haupteinzugsgebiet eines privaten Krankenhauses. Das Haus hat ausgehend davon zwei Stationen geschlossen und die Bettenzahl reduziert. Die Fallzahlen sinken. Bei einem der privaten Anbieter steigen die Fallzahlen stetig. Die Konkurrenzsituation ist sehr unangenehm. Die Frage ist: Wodurch differenzieren sich Krankenhäuser im Wettbewerb? Hierfür sind eine sehr gute Datenlage und natürlich ein PM von entscheidender Bedeutung. Beim interviewten Krankenhaus herrscht aus ökonomischer Sicht Angst vor. Dagegen hilft nur ökonomisches Denken, das im betrachteten Krankenhaus einen teils schweren Stand hat. Strukturanpassungen sind längerfristig nötig. Der Zwang zur betriebswirtschaftlichen Führung steigt. Die Dichotomie zwischen Medizin und Betriebswirtschaft ist im betrachteten Haus dominant. Die Größe des Krankenhauses hat weniger Einfluss auf die ökonomische Entwicklung als vielmehr die folgende Frage: „Woher kommen die im Krankenhaus arbeitenden Leute, in welcher Art denken sie?" Die Organisation wurde verändert und es resultierten neue Verantwortlichkeiten. Der Direktor des Krankenhauses ist kein Arzt und investierte stark in die Organisationsentwicklung. Er entwickelte eine Strategie für das Krankenhaus, die jährlich reviewt und angepasst wird. Dies bedeutete auch einen Bedeutungszuwachs für die Finanzen, die neu als Bereich in die Geschäftsleitung aufgenommen wurden. Dies hatte auch für die den Finanzen zugeordnete IT positive Auswirkungen. Zuletzt waren 29 Personen unter dem Interviewten in folgenden Bereichen tätig: Hotellerie, Finanzen, Logistik, Facilities Management sowie Administration. Die IT-Situation präsentierte sich wie folgt: Eines von zwei HIS-/ERP-Systemen wurde eliminiert. Damit trat aus IT-Sicht eine Konsolidierung ein. Neu wurden auch ein DWH-System und ein Reporting-Werkzeug (Cube-basiert) eingeführt. Die DRG-Einführung führte zur Frage: Wie steuert man Fälle (DRG's)? Die Abrechnungsthematik war organisatorisch ein viel dominanteres Thema als technisch. Die Produktivität (Effizienzsteigerung) in der Abrechnung musste gesteigert werden, Abrechnungen schneller und präziser erfolgen, was zu einer schnelleren Verrechnung führte. Die Codierung musste neu organisiert und professionalisiert werden, mit positiven Wirkungen auf Austritts- oder Arztbriefschreibung (Beschleunigung durch Prozessverbess.). Zudem musste die Informationsbereitstellung für die Zuweisenden überdacht werden. Für die Fallsteuerung hatten sich zum Schluss drei Steuerungsformen herauskristallisiert: Individualsteuerung (Cases), Behandlungspfade und retrospektive Steuerung der Cases. Individualsteuerung (Cases): Dies bedeutete, dass vor dem Eintritt bereits gesteuert wird. Präcodieren der DRG's, laufend nachcodieren, Informationen möglichst schnell bereitstellen, Leute für Monitoring einsetzen, etc. Dies war aber deutlich zu aufwendig. Nirgendwo in Deutschland und der Schweiz wurde präcodiert. Behandlungspfade: Wesentliches Thema; wie viel braucht's dafür? Was muss genau beschrieben werden? Hier stellte sich Unhandlichkeit als Hinderungsgrund heraus. Wen interessiert was? Wenn 10 bis 15 Seiten Behandlungspfadbeschreibung folgen, interessiert das Dokument niemanden. Es war schwierig einen praktikablen Mittelweg zu finden. Erforderlich waren sehr einfache Anleitungen, was bei der Verschiedenheit der medizinischen Krankenhausleistungen aber nicht einfach zu bewerkstelligen war. Die Mengensteuerung stand zunehmend im Vordergrund. Es gilt Economies of scales zu realisieren. Dabei entstand ein Problem zwischen Ärzten und Pflege. Die Pflege hat bezüglich Economies of Scale

andere Vorstellungen als Ärzte. Zwei Tendenzen waren zu beobachten: „Miteinander zusammenarbeiten“ versus „Leitung versus Mitarbeiter“. Hinsichtlich Baserates zu den DRG's ist das Krankenhaus gut aufgestellt. Es weist im Schnitt etwa 80 Prozent Kosten für den eingenommenen Preis (baserate) auf. Patienten dürfen damit nicht zu kurz aber auch nicht zu lange liegen. Das führte zu Perversionen, sodass etwa ein sehr effizientes (mikroinvasives) Operationsverfahren für Leistenbrüche bezüglich Verweildauer verlängert werden musste, damit keine Kurzlieger oder Nichtlieger resultierten. Dies hätte eine Reduktion der Einnahmen bedeutet. Retrospektive Steuerung: Hier können z.B. Ausreisser-Diskussionen u.a. mit Ärzten zu Fall-Ausreissern erwähnt werden. Das Ziel war ein Lernen für die nächsten Cases. Auch das war nicht einfach. Der Patient darf nicht unter der Layoutlier-(Kurzlieger)Untergrenze aber auch nicht darüber (Langlieger) liegen. Die DRG gibt für den Case an, welches die ideale Verweildauer ist. Vertrackt ist, dass für den Case immer der gleiche Betrag entgolten wird, was dazu führt, dass die Steuerung in Richtung der „blauen Linie“ (tiefere als durch SwissDRG festgelegte Fallkosten) gehen sollte. Im Wesentlich galt es damit die Kostenlinie unter der Baserate-Linie zu halten. Am Lernen ist man bis heute bezüglich des Verhältnisses Kosten versus Baserate. Die Codierung zu verbessern bedeutet, dass eine früher nicht codierte Leistung (z.B. Komplikation) heute codiert wird. Es kann sein, dass für den Case damit mehr Geld resultiert, weil eine neue DRG zuteilbar ist. Austrittsberichte und Arztbriefe mussten auch verbessert werden. Das bedeutete auch Ärzteschulungen, was viel Aufwand bedeutete. Ziel ist es, ausgehend von der Codierungsverbesserung den Umsatz zu erhöhen, andere DRG's einzusetzen, z.B. bei Komplikationen, und damit das Kosten-Ertragsverhältnis zu verbessern. Im Bereich OP erfolgten Verbesserungen basierend auf dem Beizug von externen Benchmarks, auch aus dem Ausland. Es erfolgte eine Analyse im Hinblick auf Strategie, Organisation und Benchmarks. Aus terminlicher Sicht und bezüglich des Einführungsdatums der DRG's ausgehend vom neuen Krankenhausfinanzierungsgesetz wurde ein integriertes Vorgehen für die organisatorischen Anpassungen und die dazu erforderlichen IT-Anpassungen implementiert. KIS- und ERP-Anbieter waren ja auch bei anderen Krankenhäusern mit der DRG-Implementierung zugange und hatten damit eine Art Vorlauf. Entsprechend wurden Änderungen bezüglich DRG-Einführung einfach nachvollzogen. Mit der organisatorischen Vorbereitung des Krankenhauses im Hinblick auf die DRG-Einführung begann man bereits 2011. Der Arztbrief musste per 1.1.2012 stehen. Das war auch das Stichdatum für die DRG-Einführung. Das ERP-System lief, für die DRG-Einführung per November 2011 schon entsprechend vorbereitet, da dann auch schon die Arztbriefinhalte bekannt waren. Die weitere Reihenfolge der Realisierung lautete wie folgt: 1. Arztbericht; 2. Codierung November 2011 produktiv; 3. Abrechnungsschnittstelle, produktiv per Januar 2012; 4. Grundabrechnung Januar, per Februar 2012 produktiv; 5. Sonstige Datenerfassung – Neue Anforderungen – Zusatzdaten erforderlich – produktiv per Januar 2012; 6. Abrechnung lückenhaft, Behebung Februar / März 2012; 7. Elektronische Datenübermittlung an Krankenkasse ungleich TARMED (tarif médical – ambulant) – Neues XML-Format für stationäre Behandlungen nötig, Rechnung – Codierdaten, etc.; 8. Im April 2012 ging die erste Rechnung raus; 9. Im Mai/Juni 2012 konnten 80% der Normalfallrechnungen elektronisch abgewickelt werden. Viel Zeit verstrich mit organisatorischen Streitigkeiten

und Nachfakturierungen; 10. Spezialfälle stellen bezüglich DRG-Abrechnung noch heute ein Problemfeld dar. Vollständig elektronisch verläuft die Abrechnung auch heute noch nicht. November 2012 blieb dann noch 1 Prozent der Fälle liegen; sehr schwierige Fälle. Die Analytik auf den Daten zu Auswertungen lief ab Juli 2012 auf einer auf Krankenhäuser spezialisierten DWH-Lösung an. Dabei handelt es sich um eine DWH-Lösung mit Cubes/Reporting-Möglichkeit, einschließlich BfS-Daten/-Statistikbereitstellung. Kostenrechnungssimulationen konnten mit dem Werkzeug angeboten werden. Ab Juni/Juli 2012 waren mittels des Werkzeugs die folgenden Resultate/Auswertungen erhältlich: Hitliste zu DRG's, Geschäftsleitungs-Reports, finanz- und leistungswirtschaftliche Auswertungen, Ressourcenauslastungen, Schnitt-Nahtzeiten sowie Entlassungszeiten. Auswertungen und Reports wurden in Zusammenarbeit mit den Ärzten laufend angepasst und stark weiter entwickelt.

2.2 Fallstudie B – Großes Krankenhaus

Im entsprechenden Krankenhaus läuft zurzeit ein Projekt Steuerung oder PM. Im Projekt arbeiten neben dem Interviewten noch zwei Personen. Vor zwei Jahren agierte man hier noch mit großen Excel-Tabellen. Im Wesentlichen geht es um die Entwicklung eines Kennzahlenmanagement(s/-baums), über das (den) gesteuert und geführt werden kann. Die Einführung von Management Cockpits ist ein wesentlicher Teil davon, doch sind solche aktuell nicht verfügbar. Deren Einführung ist abhängig von der Einführung eines neuen HIS- oder ERP-Systems. Das Thema PM ist vielschichtig. Aktuell wird das Finanzreporting eher Rechnungswesen-lastig geführt. Auch ein Arzt ist, um zu führen und zu steuern, auf Kennzahlen angewiesen. Der Treiber für das Projekt Steuerung und PM ist die DRG-Einführung. Dabei stellt sich die Frage, wofür Transparenz geschaffen werden soll. Das Krankenhaus will eine schwarze Null schreiben. Es ist nicht interessiert an Gewinnmaximierung. Allerdings müssen die internen Investitionen durch die Einnahmen finanziert werden können und es darf kein Verlust entstehen. Aktuell ist z.B. ein neuer millionenteurer Neubau geplant. Die Krankenhausleitung hat eine Krankenhaus-Strategie entwickelt, deren Umsetzung aktuell ansteht. Deren Überwachung erfolgt über Messungen in deren Zusammenhang auch immer wieder die Diskussion um die Qualität in der Medizin geführt wird. Das Krankenhaus ist aktuell im Kanton unter den Top-Drei-Häusern positioniert. Ausgehend von dieser Diskussion erläuterte der Inverviewte das Thema Baserates in Zusammenhang mit den DRG's. Die Baserate liegt aktuell im Kanton bei CHF 9500 und das Krankenhaus hat im Schnitt Fallkosten in Höhe von etwas unter CHF 8000. Die Differenz ist als Gewinn aus dem Tagesgeschäft zu betrachten und wird u.a. für die erwähnte Gebäude-Investition genutzt. Der Umsatz war im Jahr 2012 ein dreistelliger Mio.-Betrag. Das Kostenbewusstsein ist groß und steht zum Teil „buchstäblich an Türen angeschrieben", etwa in der Chirurgie. DRG's stellen somit ein Anreizsystem für die Weiterentwicklung des Krankenhauses dar, was zu einem auf Basis eines PM und auf Kennzahlen basierter Führung führt. Organisatorisch ist das Krankenhaus in Departemente und Fachbereiche gegliedert. Über den organisatorischen Wandel konnte die „Kultivierung von Gärtchen" (bei größeren Häusern fast immer der Fall) eliminiert werden. Heute steht dem Krankenhaus eine paritätisch zusammengesetzte

Geschäftsleitung vor. Darin ist die Dichotomie „Betriebswirtschaftliche Führung“ versus „Medizinische Führung“ kein Thema mehr (fruchtbare Zusammenarbeit). Ins erwähnte Change-Projekt PM, das seit ca. 18 Monaten läuft, wurden somit klar technische und Finanzressourcen miteinbezogen. Es gilt für Systemeinführungen der Primat der Anforderungen. Wenn nötig werden Architekturveränderungen ins Auge gefasst. Es erfolgte im Rahmen des Projektes kein Shift von betriebswirtschaftlicher zu medizinischer Leitung. Im Krankenhaus wird klar ein CEO-Modell gelebt. Der aktuelle CEO ist Ökonom. Dies hat Tradition beim Krankenhaus. Die Geschäftsleitung ist aktuell aus fünf Ärzten (Direktoren Departemente/Chefärzte) und fünf Nichtärzten (einschließlich des Ökonomen, der für das Interview zur Verfügung stand) zusammengesetzt. Die Ärzte verhalten sich immer mehr als Manager. Die Leitung des Krankenhauses versteht sich als Einheit. Die miteinander verbunden zu verstehenden strategischen Ziele der Geschäftsleitung des Krankenhauses lauten auf Wirtschaftlichkeit und Qualität. In Beidem ist das Krankenhaus führend in der Schweiz. Reporting-mäßig wurden früher Mega-Excels kreiert für die Führung. Geplant ist nun ein Management-Cockpit mit roten und grünen Ampeln als Hilfsinstrument zur Führung. Ziel ist es die Kennzahlen stark zu reduzieren. Die Krankenhaus-IT umfasst etwa drei Dutzend IT-Mitarbeiter, was im Verhältnis zu den Beschäftigten des gesamten Krankenhauses (tiefe vierstellige Zahl) relativ klein ist. Die IT hat eine geringe Kapazität und ist sehr stark ausgelastet/überlastet mit dem täglichen Geschäft. Wie erwähnt steht aktuell eine HIS- oder ERP-Einführung an, die mit der Problematik der Business-Intelligence und PM-Thematik verknüpft betrachtet wird. Die BI-Anwendung soll nicht selber entwickelt werden. Jedoch ist dafür Know-how-Aufbau nötig (zusätzliche vier Vollzeitäquivalente für Betrieb und Anwendungspflege). In der Vergangenheit wurde zu viel selber entwickelt.

2.3 Fallstudie C – Sehr großes Krankenhaus

Das sehr große Krankenhaus macht einen sehr hohen Umsatz, u.a. auch mit Einnahmen aus dem Ausland. Es arbeitet eine mittlere vierstellige Zahl an Mitarbeitern im Krankenhaus. Das Management der Finanzen ist als sehr herausfordernd und mächtig zu bezeichnen. Mit betriebswirtschaftlichen Fragen/PM ist das Krankenhaus stark beschäftigt, auch ohne DRG-Einführung. Der Direktor, erstmals eine ökonomisch ausgebildete Person, legt Wert auf stärker betriebswirtschaftliche Führungsmotive: Führen mit Zahlen. Der Interviewte arbeitet seit ca. einem halben Dutzend Jahren im entsprechenden Krankenhaus. Das PM wurde im Krankenhaus bereits seit anfangs Zweitausender-Jahre aufgebaut. Damals operierte man mit Zahlenreihen, z.B. auf Basis medizinischer Daten. Die finanzielle Führung spielte damals keine so große Rolle. Es wurde eher politisch geführt und nicht auf Basis finanzieller Kennzahlen. Das ist heute anders. Ab Anfang der 2000er-Jahre wurde auch ein großes ERP-System eingeführt. Damit begann das Zeitalter des Management Accounting. Eine Kostenarten-, -stellen- und -trägerrechnung wurden eingeführt. Circa nach der Mitte der Nuller-Jahre wurde damit begonnen, Reporting und Analysen zu systematisieren. Zu betonen ist hier, dass das Krankenhaus bei der Menge der Reporting-Adressaten, die zu bedienen sind, nie gewillt war, ein individuelles Reporting zuzulassen. Zentral sind beim Zahlen-orientierten Reporting ausführliche

Kommentarfelder, die dezentral (durch medizinische Abteilungen) oder zentral bei den Finanzen „befüllt“ werden. Dabei setzt man auf ein Kennzahlensystem von ca. 30 bis 35 Stück, was reicht. Ebenfalls werden ca. 30 bis 35 Standardberichte eingesetzt. Unterdessen ist auf Basis eines BI-Tools eine Führungscockpit-Lösung in Entwicklung, die in den folgenden Monaten frei gegeben wird. Die Geschäftsleitung (GL) sagt wohin die Reise geht. Die Arbeit ist nicht immer einfach, weil vielfach stark politisiert. Die Abt. Finanzen baut das Cockpit zu den GL-Vorgaben. Zwischen der Geschäftsleitung/-Direktionen und den Medizinbereichen (Kliniken) besteht eine gewisse Distanz. Dies führt dazu, dass die „Führung der medizinischen Direktionen oder Institute“ hauptsächlich mit den angesprochenen Kennzahlen(rastern) erfolgt. Das Zusammenwachsen der verschiedenen Direktionen/Medizinbereiche erfolgt nur langsam. Für diese positive Entwicklung ist auch der Spardruck ausschlaggebend. Laut Aussagen des Interviewten lernt man (etwa im Bereich PM) viel. Das bestehende DWH-System dient als Grundlage für die Auswertungen. Die Cubes werden durch eine eigens designte Lösung abgelöst. Bei der Usability achtete man besonders auf Management-Tauglichkeit (Miteinbezug einer Designerin in GUI-Entwicklung). Der Cockpit-Bildschirm (browser-basiert, auch mobil nutzbar) unterscheidet in einer Matrix Kennzahlen und Direktionen/Medizinbereiche, innerhalb der mit einfachen farblichen und symbolhaften Ampeldarstellungen dargestellt wird, welche Ausprägung die Kennzahlen haben (grün, rot, etc.). In alle Richtungen sind Drill-downs auf Detaildaten möglich. Aktuell gibt es sogenannte Trimester-Berichte (in Papier). Dreimal jährlich erhalten Geschäftsleitung (Direktionsmitglieder) und Krankenhaus- oder Verwaltungsrat einen „Stapel Papier“, der systematisch von der Gesamtübersicht über Direktionen und Klinikdetails differenziert; immer in sogenannten One-Pagers, mit Kommentarspalten seitens Finanzen oder für die dezentralen Stellen. Von den Trimesterberichten ist unterdessen ca. 80% standardisiert generierbar, der Rest wird manuell gepflegt. Bei den Klinikberichten werden Gegenkommentare dort von der Finanzabteilung eingetragen, wo diese mit den Angaben der Kliniken nicht einverstanden sind. Pro Organisationseinheit wird ein Onepager erstellt. Aus den Trimesterberichten resultieren Beschlüsse. Bei der IT ist die Finanzabteilung ein Anforderungssteller. Die Finanzen haben keine eigene IT, genauso wie alle anderen Abteilungen, Direktionen sowie Kliniken. In der Geschäftsleitung des Krankenhauses sind die Finanzen berücksichtigt, nicht aber HR, was eher unüblich ist. Mitberücksichtigt in der GL ist auch die IT (CIO). Dies führte in den in der Direktion vertretenen Bereichen klar zu besonderen Entwicklungen. Die IT ist laut Aussage des Interviewpartners sehr nahe am Geschäft dran. Manchmal würde man sich aus Geschäftssicht seitens IT etwas mehr Tempo wünschen. In den Bereichen Finanzen, Betrieb und IT ist in der Vergangenheit großer Nachholbedarf vorhanden gewesen, deshalb hat man die GL entsprechend konfiguriert. Das Krankenhaus macht im Hinblick auf das PM rasche Fortschritte. Im Vergleich ist es für die Schweiz ein sehr teures Krankenhaus. Das hängt mit der Zentrumsfunktion, der Vereinigung von Forschung und angewandter Medizin, aber auch mit der Betreuung vieler teurer (Spezial-)Fälle zusammen (DRG-Problematik). Die Methodik der (Cockpit-)Berichterstattung hat sich unterdessen geändert. Es werden in einer Pyramide drei Hierarchieebenen unterschieden. Aus der Pyramide werden „drei Schnitze“ mit unterschiedlichem thematischem Fokus herausgeschnitten: Leistungen/Finanzen, Produkti-

vität/Auslastung sowie Prozesse (in Entwicklung). Das Cockpit-Werkzeug basiert auf drei Bereichen. Es sind Quadratgrafiken mit Positionierungen der eigenen Organisationseinheit in vier Quadranten möglich, wo die Organisationseinheit sehen kann, ob sie sich gegenüber dem vergangenen Zeitabschnitt verbessert oder verschlechtert hat (Positionierung durch unterschiedlich große Kreise). Drei Bereiche können darin unterschieden werden: Dashboard, Topics, Compare Tool (Ranglisten-Bildungen). Letzteres ermöglicht Benchmarks über die verschiedenen Kliniken und Direktionen. Letzteres ist für die Mediziner sehr wichtig. Vergleiche sind etwa über Benchmarks oder auch über Deckungsbeitrag oder Gewinn machbar. Transparenz ist der Primat der Stunde: Daher geht die Philosophie immer mehr in Richtung Cockpits einer Organisationseinheit oder Klinik auch für andere Kliniken einsehbar zu machen. Nicht nur DRG's sind Treiber der Krankenhausentwicklung. Der politische Führungsimpuls ist weniger dominant als zunehmend das unternehmerische Handeln. Zwischen öffentlichen und privaten Krankenhäusern besteht keine Differenz mehr mit der neuen Krankenhausfinanzierung und den entsprechenden Gesetzen. Wachstum, Gewinn, Opportunitäten, Markt sind die Treiber. Das Krankenhaus generiert denn auch mit neuen Leistungen immer mehr Drittumsatz und gute Margen. Gegenwärtig werden etwas weniger als drei Dutzend Mio. CHF Drittumsatz gemacht. Tendenz steigend. Im Moment geht deshalb eine Gratwanderung ab aus Kostenreduktion und Wachstum. Die DRG-Einführung stellt nichts anderes als die Einführung eines neuen Preissystems dar. Auch das betrachtete Krankenhaus finanziert seine Investitionen selbst. Gewinne wecken politisch Begehrlichkeiten. Krankenkassen (KK) und die Politik fordern dann Einsparungen, respektive Abgaben, was aber nicht sein dürfte, da das Krankenhaus ja wie erwähnt auch seine Investitionen selbst decken muss. Das Zeitalter der schwarzen Null, die früher Pflicht war, ist vorbei. Gewinne kann man daher gegen außen am ehesten dadurch verkaufen, dass man in dem Zusammenhang damit auch gleich Sparprogramme verkauft und kommuniziert. Wichtig sind aus Sicht des Steuerungssystems Krankenhaus die Anreizsysteme, die müssen gut sein. Das Krankenhaus fällt wegen seiner Größe teilweise aus dem Raster der Krankenhausfinanzierung.

3 Tabellarische Hypothesenprüfung

Im Folgenden wird zusammenfassend angezeigt, welche der weiter oben formulierten Hypothesen angenommen werden können und welche abgelehnt werden müssen. Die Auswertung der Hypothesen aus den Fallstudien zeigt klar, dass die Hypothesen 1 und 2 (partiell), 4 (partiell), 5 und 7 vollständig und über alle Krankenhäuser hinweg angenommen werden können. Die Hypothesen 3 und 6 sind abzulehnen. Die Annahmen lassen folgende Schlussfolgerungen zu. Hypothese 1: Organisatorische Maßnahmen stehen für Krankenhäuser klar im Vordergrund für das PM und nicht Technologieimplementierungen. Hypothese 4: Die Maturität der IT in Schweizer Krankenhäusern ist relativ gering. Nur in einem von drei Krankenhäusern sitzt der CIO in der Geschäftsleitung. Die geringe Maturität führt dazu, dass keine sehr sophistizierten Auswertungsmechanismen bereitgestellt respektive eingesetzt werden. Im Wesentlichen werden Repor-

tings für die Steuerung der Krankenhäuser vorgegeben oder ist Ähnliches intendiert. Es werden aktuell Cube-/Reportinglösungen in zwei von den drei befragten Krankenhäusern eingesetzt. Hypothese 5: Wenn die organisatorische Veränderung der Krankenhausorganisation von einer Expertenorganisation hin zu einer betriebswirtschaftlich geführten Organisation führen soll, dann müssen entweder die medizinischen Experten zu stärkerem betriebswirtschaftlichem Denken angehalten werden oder aber die betriebswirtschaftlich orientierten Stellen im Krankenhaus müssen bezüglich Führungsfunktionen im Krankenhaus aufgewertet werden. In zwei von drei Krankenhäusern hat die medizinische Führung Vorrang vor der betriebswirtschaftlichen Führung, was in einem Fall (größtes Krankenhaus) abhängig ist von der betrachteten Hierarchieebene. Im Übrigen trifft die Hypothese wohl für alle Krankenhäuser zu. Im Fall des kleinsten betrachteten Krankenhauses werden klar und explizit nur 10% bis 20% des IT-Budgets in betriebswirtschaftliche Lösungen investiert. Im mittelgroßen Krankenhaus erfolgen die BI- und DWH-relevanten Investitionen erst noch, wenngleich schon Cockpit-Lösungen im Einsatz sind. Bei den meisten Krankenhäusern dominiert damit die medizinische Führung zumeist klar. Die Hypothese 6 ist klar abzulehnen. Eine Selbstüberschätzung tritt in den Fallstudien seitens IT nirgendwo auf. Einerseits weil die IT fast überall chronisch mit dem Tagesgeschäft überlastet ist, andererseits weil zum Teil harmonisch Überarbeitungen bestehender Architekturen (Ersatz und Eliminierung redundanter IT-Systeme, Aufbau oder Weiterentwicklung von DWH-Systemen und BI-Werkzeugen) sich „sanft“ in die DRG-Einführung einbetten. Die Hypothese 7 kann klar angenommen werden. Betriebswirtschaftliches Know-how ist überall erforderlich und die IT muss zusehen, wie sie den Steuerungs- und PM-Impetus bestmöglich unterstützen kann. Weniger explizit wird an der IT-Maturität gearbeitet. Expliziter wird an der betriebswirtschaftlichen Maturität gearbeitet; sei es, dass, wie im kleinsten Krankenhaus, explizit Steuerungsmöglichkeiten fachübergreifend intensiv diskutiert werden, sei es, dass Großprojekte für das PM initiiert werden (mittelgroßes Krankenhaus), sei es, dass wie beim größten Krankenhaus intensive Schulungen und bidirektionale Reporting-Runden resultieren.

Nr. der Hypothese	Kleines Krankenhaus	Großes Krankenhaus	Sehr großes Krankenhaus	Hypothesen-Prüfung
1	+++	+++	+++	Annahme
2	++	+++	+++	Annahme
3	---	---	---	Ablehnung
4	+++	+++	---	Annahme
5	+++	+++	+++	Annahme
6	---	---	---	Ablehnung
7	++	+++	+++	Annahme

Abbildung 3: Tabellarische Hypothesenprüfung.

4 Tabellarische und verbale Schlussfolgerungen zu den Fallstudien

Die folgenden Schlussfolgerungen können in tabellarischer Form zum Fallstudienvergleich gezogen werden (Die relativ große Tabelle 1 zum differenzierten Vergleich der

Fallstudien ist aus Platzgründen im Dokument zu finden, das via Dropbox-Ordner aufgerufen werden kann, dessen URL im Anhang genannt wird.). Unterschiedliche Ausprägungen artikulieren sich ausgehend von der Größe des Krankenhauses. Ursprünglicher Gedanke: Je größer das Krankenhaus, desto expliziter oder stärker ausgeprägt ist die Dichotomie zwischen Betriebswirtschaft und Medizin auch organisatorisch. Dies trifft nicht ganz zu. Im kleinsten Krankenhaus steht die Dichotomie auch im Vordergrund, allerdings wurden im Rahmen der DRG-Einführung doch erhebliche Alignment-Anstrengungen durch die intensive Diskussion von Steuerungsmöglichkeiten unternommen. Im mittelgroßen bis großen Haus hingegen wurde einerseits über ein großes PM-Projekt das Thema sehr direkt adressiert. Zusätzlich arbeitet die Geschäftsleitung konsensorientiert und ohne trennende Dichotomie zwischen medizinischer und betriebswirtschaftlicher Führung. Im größten betrachteten Krankenhaus herrscht nicht nur auf Geschäftsleitungsebene eine teilweise Dichotomie zwischen medizinischer und betriebswirtschaftlicher Führung vor, sondern aufgrund der Größe auch zwischen Geschäftsleitung und den dezentralen klinischen Leitungen. Hier bleibt nichts anderes übrig, als dass sehr stark kennzahlenorientiert geführt wird (Reporting-Orientierung), um eine einheitliche Sicht über die Performance der Gesamtorganisation zu erlangen. Damit können auch Muster im Sinne weiterer Hypothesen formuliert werden. Je kleiner das Krankenhaus, desto größer die Wahrscheinlichkeit, dass die medizinische Leitung klar dominierend ist. Je größer indessen der Wettbewerbsdruck auf das kleine Krankenhaus ist, desto größer ist die Wahrscheinlichkeit, dass auch das medizinische Personal das PM aufnimmt. Dies ist schon rein der langfristigen Überlebensfähigkeit geschuldet. Sehr große Krankenhäuser haben erhebliche Schwierigkeiten eine betriebswirtschaftliche Steuerung klar in den Vordergrund zu stellen, da sehr spezialisierte und ausdifferenzierte medizinische Domänen klar im Vordergrund stehen. Überdies nimmt auch die Komplexität der IT alleine der Größe wegen zu, u.a. wegen des Mixes aus dezentral und zentral geführter IT. Ausgehend von der geringen Maturität der IT (in sehr großen aber auch in kleineren Häusern) nimmt das Problem IT für das PM künftig eine immer dominantere Rolle ein, auch die dafür erforderliche Datenintegration. Für ein flexibles PM in Krankenhäusern spielt die Flexibilität der IT eine zentrale Rolle. Alle Krankenhäuser haben einfache Reporting- und OLAP-Funktionalität integriert. Jedoch kommt noch selten individuelle Self-Service-Auswertung dazu. Im Fall des größten Krankenhauses wurde dies sogar explizit nicht gewählt, weil ansonsten das Management der entsprechenden Applikation aus Sicht der Finanzen als „Applikationseignerin“ zu komplex würde.

Literaturverzeichnis

[Af13] Affolter, C. (2013): Neue Spitalfinanzierung – Spannungsfeld zwischen Markt und Regulierung. http://www.swissdrg.org/assets/pdf/Forum2013/7_SwissDRG_Forum_santesuisse.pdf (Aufruf per 2013-08-07).

[BAG13] BAG (2013): SwissDRG. http://www.bag.admin.ch/themen/krankenversicherung/06492/12987/ (Aufruf per 2015-04-26; erstellt per 2013-07-09).

[CCT11] Campanale, C.; Cinquini, L.; Tenucci, A. (2011), Do management accounting systems influence organizational change or vice-versa? Evidence from a case of constructive research in the Healthcare Sector, 34 EEA Annual Congress.

[Ch84] Child, J. (1984): New technology and developments in management organization. In: Omega 12, 3, S. 211-223.

[DC99] Davidson, E.J.; Chismar, W. (1999): Examining the Organizational Implications of IT Use in Hospital-based Health Care: A Case Study of Computerized Order Entry. http://www.phdmanagement.sssup.it/documenti/awarded/campanale_thesis.pdf (Aufruf per 2013-01-27; erstellt 1999).

[HF11] Herzog, J.; Funke, C. (2011), Reifegrade im Corporate Performance Management, Zürich: KPMG.

[Me09] Mertens P. (2009): Integrierte Informationsverarbeitung – 1. Operative Systeme in der Industrie. Gabler, Wiesbaden.

[MM09] Mertens, P.; Meier, M.C. (2009): Integrierte Informationsverarbeitung – 2. Planungs- und Kontrollsysteme in der Industrie. Gabler, Wiesbaden.

[NZZ11] NZZ (2011): Kantone, Spitäler und Versicherer in der Pflicht. http://www.nzz.ch/aktuell/schweiz/kantone-spitaeler-und-versicherer-in-der-pflicht-1.13390861

[Og12] Oggier, W. (2012): Privatspitäler als Gewinner der neuen Spitalfinanzierung. https://www.yumpu.com/de/document/view/32583311/download-artikel-swiss-leading-hospitals (Aufruf per 2015-06-18).

[WMH13] Walser, K.; Meister, A.; Huber, E. (2013): Performance Management: Die Spitäler haben das Problem erkannt, aber es gibt noch einiges zu tun – SwissDRG stellt hohe Anforderungen. In: clinicum (2013) 2. S. 1-6.

[WHMoJ] Walser, K.; Huber, E.; Meister, A. (ohne Jahr): Planning, budgeting and performance management at Swiss hospitals – Are Swiss hospitals at a crossroads – will these medical specialist organisations in future develop into institutions with a business orientation?. Erscheint in MIBE (http://www.egms.de/static/de/journals/mibe/)

[Yi09] Yin, R.K. (2009): Case Study Research – Design and Methods, Sage, Los Angeles et al.

Anhang

Die in Kapitel 4 erwähnte Tabelle zum Fallstudienvergleich sowie der Interviewleitfaden können unter dem folgendem Dropbox-Link als PDF-Dokument aufgerufen werden:

https://www.dropbox.com/s/s6nnexjmtf1l8zc/Dropbox_Auslagerungen_HospPerf_CaseStud%20III%20-%20Kopie.pdf?dl=0

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

System Analysis of Information Management

Christian Kücherer [1], Manuel Jung [2], Franziska Jahn [3], Michael Schaaf [3], Kais Tahar [3], Barbara Paech [1], Alfred Winter [3]

Abstract: The successful delivery of IT services is enabled by information management that can be distinguished into strategic, tactic and operational information management. A major challenge of information management is the structuring and definition of its tasks, entity types and tools. At the same time, the increasing complexity of IT services and related infrastructure, cost effectiveness and increased users expectations demand standardized processes for operational information management. One major approach to structure and standardize IT operations is the ITIL©[4] framework, the Information Technology Infrastructure Library. ITIL contains best practices to align IT services with the needs of a business in general. Several organizations have introduced processes of ITIL in their operational information management.

As an example of an organization, we study a hospital. So far there is no agreed standard of what constitutes proper information management in a hospital. Therefore it is interesting to study individual hospitals and their information management. This paper describes the results of a system analysis based on 7 interviews to investigate a hospital's information management tasks, entity types, tools and the adoption of ITIL in order to analyze strengths and weaknesses of IT support of Information Management.

Keywords: Information Management, System Analysis, ITIL

1 Introduction

According to Winter et al. [Wi11], information systems (IS) play an important role for many people within organizations in general and especially in organizations of the health care environment. Staff of health care institutions are affected by the availability and quality of information in their daily work, especially in hospitals. For the efficient operation of IS, an information management (IM) is necessary, which is separated into strategic, tactical and operational information management.

Strategic IM establishes strategies and principles for the information management division that result in a strategic IM plan. Tactical IM follows the needs of strategic IM and implements solutions or changes within projects. The provisioning of IT services is the primary task of operational IM.

[1] Universität Heidelberg, Institut für Informatik, Im Neuenheimer Feld 326, 69121 Heidelberg, (kuecherer | paech)@informatik.uni-heidelberg.de
[2] Universität Heidelberg, Im Neuenheimer Feld 326, 69121 Heidelberg, manuel.jung@stud.uni-heidelberg.de
[3] Universität Leipzig, Institut für Medizinische Informatik, Statistik und Epidemiologie, Härtelstr. 16-18, 04107 Leipzig, (franziska.jahn | michael.schaaf | kais.tahar | alfred.winter)@imise.uni-leipzig.de
[4] ITIL©is a registered trademark of Office of Government Commerce. Further occurrences of ITIL©relates to this registered trademark.

All activities of IM consist of tasks that use or update information and are supported by tools in organizations. E.g. the task `strategic IM planning` is responsible for aligning the IM division's goals with the goals of the hospital. The entity type used to document this information is the `strategic IM plan`. Several tools might be involved in the creation of this plan. This could be an `office system` in combination with a `enterprise architecture management tool` or a `balanced scorecard tool`.

Some of the tasks, especially tasks of operational IM have adopted best practices such as ITIL. ITIL, the Information Technology Infrastructure Library [Ar13], [Ra11], [St13] is one of the major approaches to structure and standardize IT operations. ITIL contains a set of best practices for service design, service transition and service operation, governed by a service strategy, allowing a continuous service improvement. Several hospitals have adopted ITIL within their operational IM [Ho11]. However, the tasks, entity types, tools and adoption of ITIL depend on the hospital needs, organizational structure and complexity of IT used. In our research we are interested to understand the IM of different hospitals in order to elicit the IT-related needs of the IM personnel, in particular of the chief information officer (CIO). As a first step we analyzed the as-is-state of the IM of a big hospital. We call this hospital in the following hospital X. It has more than 3500 employees and more than 1200 beds.

This as-is-state can be modeled using the Three-layer Graph-based Meta Model 3LGM^2 (see section 2.3). Previous publications have shown, that 3LGM^2 is suited to model tasks, entity types and tools within hospitals [Ja09] and [Wi07]. Here we use this model for the first time to analyze the IM and its adoption of ITIL.

There are two benefits from system analysis and modeling for organizations: First, the model is the base for comparability of IM between different organizations. The model unveils the ITIL conformity of processes, which can be used to determine the level of maturity, e.g. as described in [WWB15]. Second, the modeling of tasks indicates missing or unnecessary tasks and is the basis for future optimization and consolidation of the IM and it's processes.

The main contributions of this article are:

1. A description of our system analysis to identify tasks, entity types and tools within one particular hospital's IM.
2. First results of this system analysis in terms of a 3LGM^2 model
3. A first assessment, to which extent the identified tasks are based on ITIL.

The remaining part of this paper is structured as follows. We first describe the background and related work. Section 3 explains our system analysis method used to identify tasks, entity types and tools in hospital. Section 4 presents the results of the system analysis in the hospital. Finally, ideas of future work are shown and a conclusion is drawn in section 5.

2 Background and Related Work

This section defines the terms task, entity type and tool. It then describes the method system analysis by Ammenwerth et al. and introduces the characteristics and usage of $3LGM^2$ models. Finally, the concepts of an ITIL reference model are described.

2.1 Task, Entity Type and Tool

According to TORE (Task- and Object-Oriented Requirements Engineering) [PK04] tasks are executed by users in the enterprise. They are identified by considering the current activities of the users. This can be done by interviewing the user and by monitoring of workflows. The applied systems and tools in this case support the processing of tasks. These must be explicitly described and placed in the context of the work environment. Each task is described by a task description, based on the template of TORE.

Figure 1 shows the template, as an example applied to describe the task change management. The description captures aims and possible sub-tasks, responsible persons and input and output data for the representation of the entity types, as well as meta data such as the ID, date and version. With respect to ITIL the tasks represent the (sub) tasks for the ITIL processes.

Task: Change Management **ID:** A-002 **Version:** 1.2 **Date:** 30.03.2015	
Objective	Collection of changes
Subtask of	n/a
Possibility to intervene	edit request, estimation of change
Trigger	user's request
Priority	low / ~~high~~ / ~~middle~~
Initial situation	completed request form, eligible change rationale
Information In	change request form
Information Out	change report, change appraisal, changed component/system
Resources	staff of operational IM, change advisory board, division manager, possibly department manager
Application Systems / Tools	Microsoft Office (Word, Outlook), Microsoft SharePoint
Interfaces	Exchange server, SharePoint server

Fig. 1: Task template accoring to TORE [PK04], used to describe the task change management

An entity type is any kind of information that is consumed, produced or modified by a task. These can be documents, data structures, objects, files, specified by data types or even paper based artifacts. To know all entity types has two advantages. First, it makes explicit common data interfaces that enable interoperability between IS. And second, it shows which information is necessary to perform the given tasks. In consequence, this

knowledge can be used to model which information is needed by whom in order to fulfill which task.

Tools in this context are all kinds of applications, software systems or software components that support the execution of tasks, in other words any kind of IS.

2.2 System Analysis as method

The purpose of the system analysis method according to Ammenwerth et al. [Am15] is to provide an exact description of the properties of an existing IS or a set of its parts. The analysis results in the description of an as-is-state. In cases when strengths and weaknesses are important, the analysis is followed by an assessment. The method of system analysis is domain independent, yet proposed in the context of health care environments. It comprises a system analysis process model that executes analysis activities within multiple iterations. This allows rework, adaptation and improvement of the activities' results in each iteration.

Several analysis types can be used for various objectives. Beside others, the **task analysis** considers business tasks and activities in a division. The focus of **equipment analysis** is the assessment of tools and application systems for execution of tasks and associated roles. Business processes within a division are investigated during the **process and activity analysis**.

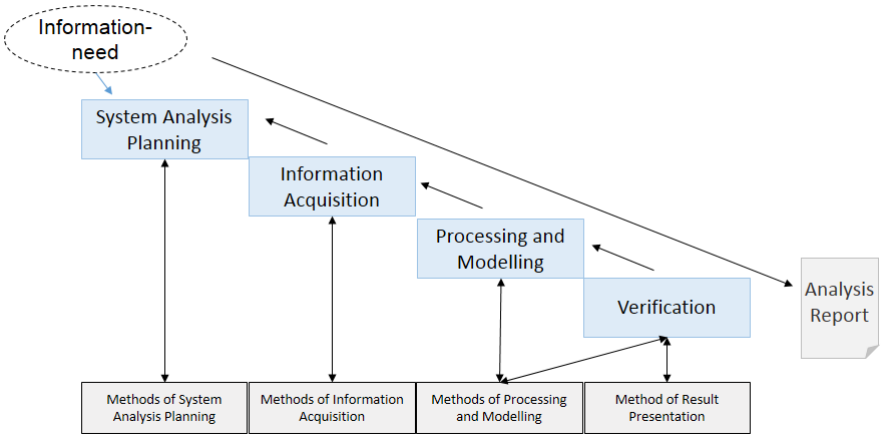


Fig. 2: System Analysis as described in [Am15]

Figure 2 shows the process model of the system analysis. A system analysis is triggered by certain objectives or a primary question. Each of the four phases contributes to the system analysis and utilizes several methods.

The first phase is the **planning of the system analysis**. It results in a documentation of the system analysis objectives as well as primary questions and the scope. Part of the planning is the definition of assessment criteria for a later assessment, if applicable. All information

needed to answer the analysis questions are identified and gathered in the second phase, the **information acquisition**. Information can be acquired by performing interviews, written surveys, observation and data analysis.

Within the third phase, **processing and modeling**, the information is processed and transformed into a suitable model. The system analysis supports any kind of model such as UML diagrams, business process models, as well as ARIS and 3LGM^2 models. The latter is discussed in detail in section 2.3.

The **verification** of the systems analysis results is done in phase four. The focus is on determination of correctness, completeness and appropriateness of the created models. Depending on the complexity of the system analysis, the verification can be either by a direct feedback with responsible persons or by verification meetings.

2.3 Modeling with 3LGM^2

The Three-layer Graph-based Meta Model (3LGM^2) has been proposed by Winter et al. [WBW03, We03]. It enables the modeling of health information systems [Wi11].

The model consists of three layers: First, the functional layer, that shows tasks and related entity types. Second, the logical tool layer that relates application systems to tasks and entity types. Third, within the physical tool layer, the tools or application systems are allocated to specific hardware, computing infrastructure and communication technology.

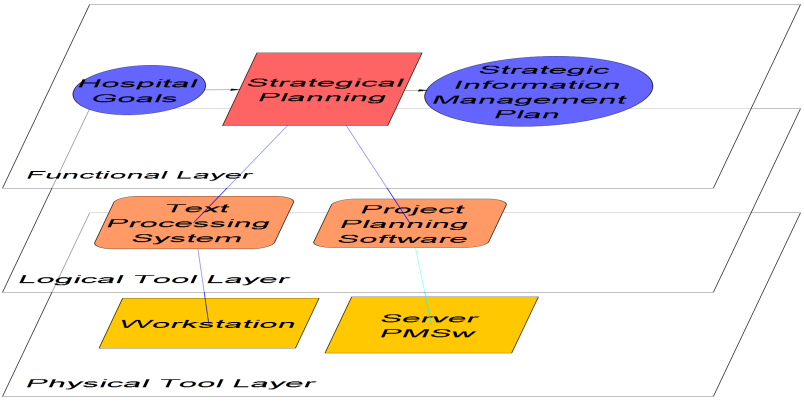


Fig. 3: An example instance of a 3LGM^2 model

Figure 3 shows a simple 3LGM^2 Model. In the functional layer, the task `strategic planning` is depicted, that uses the `hospital goals` as input and outputs a modified `strategic plan`, both entity types. The task is supported by two tools, shown on the logical tool layer: a `text processing system`, which runs physically on a user's `workstation`, and a `project planning software` that runs on a server called `PMSW`.

2.4 ITIL Reference Model

A reference model describes a class or set of models with similar properties. Such models contain the abstracted and simplified information of a domain or topic and can be used for documentation, presentation or analysis. Only through its application and usage, a model becomes a reference model [Is08]. One major feature of a reference model is its comparability with other models of a specific domain.

A 3LGM^2 based reference model for ITIL has been defined by Ißler [Is08]. The reference model condenses the application of ITIL in a medical environment and contains 9 of the ITIL v3 (published in 2007) processes. These are availability management, capacity management, incident management, IT service continuity management, release and deployment management, service asset and configuration management and service level management. On the functional layer the model contains tasks and entity types. The logical and physical tool layer do not contain any elements, since this is subject to instantiation of the reference model. All of these processes are available as independent perspectives that are partial models.

Figure 4 shows the functional layer of the ITIL process **Change Management**. There are 4 major tasks which handle the request for a change, the schedule for the change, the change and the appraisal for the change, respectively. The figure shows for each major task the three subtasks record, change and delete. In the following only the core task is explained. The first task is to handle `information about a request for change` that creates or modifies the `request for a Change`. This request is input to the task to handle `information about a schedule for a change` that results in `schedule for a change`. The schedule is used by the task to handle `information about an appraisal of a change` that creates or modifies the `appraisal of a change`. The schedule is also used by the task to `information about a change` that creates or modifies a `change`.

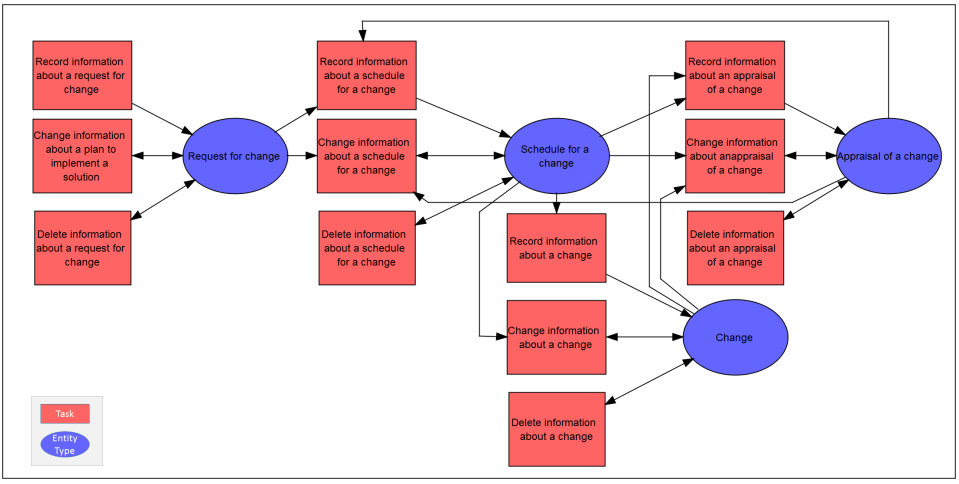


Fig. 4: Functional layer of ITIL change management [Is08]

3 Method

We have chosen the method of system analysis according to Ammenwerth et al. [Am15], since this method investigates the as-is state, addresses the health care environment and is related to 3LGM^2 . This section describes our instantiation of this method for the specific hospital X. All adaptations of the methods are explicitly described.

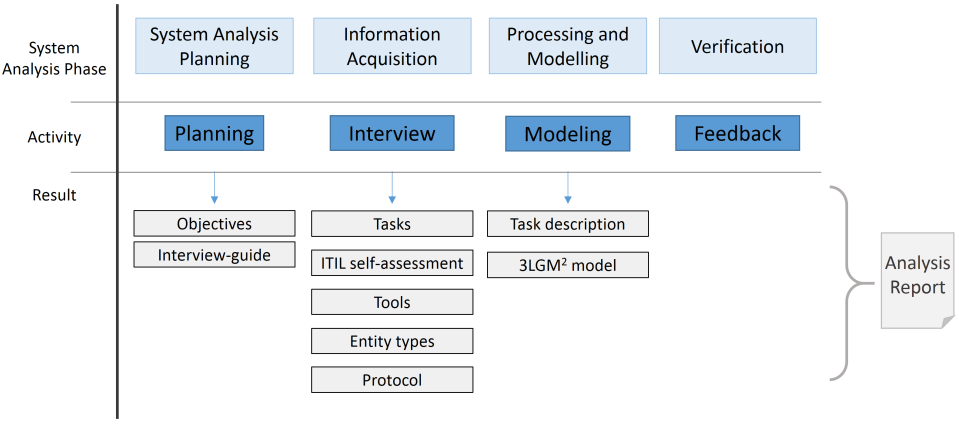


Fig. 5: The information artifacts used and produced by the system analysis

Figure 5 shows the artifacts generated in each phase of the system analysis. They are explained in each of the following subsections. All results are condensed and compiled into an analysis report.

3.1 Phase 1: System Analysis Planning

The objectives (O) of the analysis were:

O1: Elicit the as-is-state of the tasks, their related entity types and supportive tools within the IM of hospital X.

O2: Construct a 3LGM^2 model, indicating the coverage with ITIL processes.

O3: Identify any interoperability issues or media breaks within the tools.

As shown in Section 2.2, the system analysis supports different types. A task analysis discovers the IM's tasks and their relation to the ITIL reference model. Since tasks are performed within processes, the process and activity analysis identifies input and output entity types. Within the execution of tasks, tools can be identified by the equipment analysis. The system environment, organizational structures, communication paths and the content of used documents and forms was not subject of the analysis, because they do not contribute to the given objectives. Thus, we have chosen a combination of task analysis, equipment analysis and process and activity analysis.

3.2 Phase 2: Information Acquisition

The selection of an appropriate information acquisition technique (see 2.2) is based on following consideration: Observations are not appropriate due to staffs time limitation and sensitiveness of the IM division. Data analysis does not fit, since there is no explicit data available about tasks, entity types and tools used. Written surveys would work to gather an exact description of tasks, but address larger user-groups. Interviews have been found to provide appropriate results in combination with reasonable time investment and are able to gather interoperability issues in daily work.

In a first conversation with the CIO, we identified 7 persons within the IM division, who are responsible for tasks of the IM and therefore are relevant as interviewees. Non of the interviewees are related to one of the authors of this article. Interviews of several roles allow a comprehensive view of the IM's tasks, entity types and tools. The roles include the CIO, several department managers, a project manager and assistants as shown in Table 1. We conducted one interview per person, associated with one single role and used the same interview-guide for all interviews. Five interviews were face to face within a four weeks time-frame and two have been held by phone.

Interviewee	**Position/Role**	**Type**
Person 1	CIO	face to face
Person 2	Project Manager / IT Strategy	face to face
Person 3	Assistant IM	face to face
Person 4	Application Manager	face to face
Person 5	Service Manager IM	face to face
Person 6	System Manager IM	phone
Person 7	Assistant IM	phone

Tab. 1: Interviewees in hospital X

The questions presented in Table 2 were used within the interview. The table also shows the relations of the questions to the objectives and the envisioned results.

Appointments for interviews have been scheduled 2-4 weeks in advance. During the interview, one interviewer was present with the interviewee and up to two additional persons recorded the answers. Also, the interviewer were taking notes to write a protocol. The interviews had a maximum duration of 1.5 hours.

3.3 Phase 3: Processing and Modeling

Immediately after the interview, the protocols of the minute taker(s) have been combined with the interviewer's protocol into one. The obtained information regarding tasks has been transferred and assembled into the task description (see Figure 1). This task description is the basis for the 3LGM^2 modeling that creates the elements task and entity type on the functional layer. Tasks and entity types that were found to be identical for both the hospital and the reference model have been imported from the reference model into the

Goal	Question	Metric
O1	Which tasks are within your responsibility?	List of tasks
	Who else is in charge of this task?	List of contact persons
	Are there any more contact persons available for further questions?	List of contact persons
O1, O3	Which input and output information is processed?	List of entity types
	Which interfaces are used and based on what communication?	List of interfaces
O1, O2	Who is in charge of this process?	List of responsible persons
O1, O2	What ITIL-Processes are implemented in your division?	List of tasks, List of ITIL tasks
O1, O2	Which application systems are used to support each task?	List of tools
O1, O2	Are there any further application systems that are related to ITIL?	List of tools

Tab. 2: Questions in the interview

hospitals 3LGM^2 model. These elements are marked as green and have an identical ID and description, according to the elements of the reference model. The system analysis results in a hospital specific 3LGM^2 model of information management.

3.4 Phase 4: Verification

The condensed protocol has been sent by e-mail to the interviewee with the request to read the text and identify any mismatches or misunderstandings. The protocol did, however, not contain the task-description and the 3LGM^2 Model. Interviewee Person 1 responded a PDF-scan with some 30 comments written feedback. A PDF with 6 inserted comments was responded by Interviewee Person 3. A PDF with 12 inserted comments was responded by Interviewee Person 7. Their feedback has been integrated into a new version of the protocol and included into the modeling. Interviewee Person 4 responded by e-mail, that the content is alright and topics a represented correctly. No other feedback or comments were retrieved.

Once all tasks, entity types and tools are updated in the 3LGM^2 model, it will be presented to all interviewees for further consolidation.

4 Results

We have identified 13 top-level tasks and 24 tasks altogether that are supported by 47 different tools. The tasks can be hierarchical. For instance, the task change management consist of several sub tasks. For reasons of clarity, we name top-level tasks as processes that consist of several tasks. An overview of the top-level tasks is shown in Figure 6. We associated each top-level task with one level of the IM.

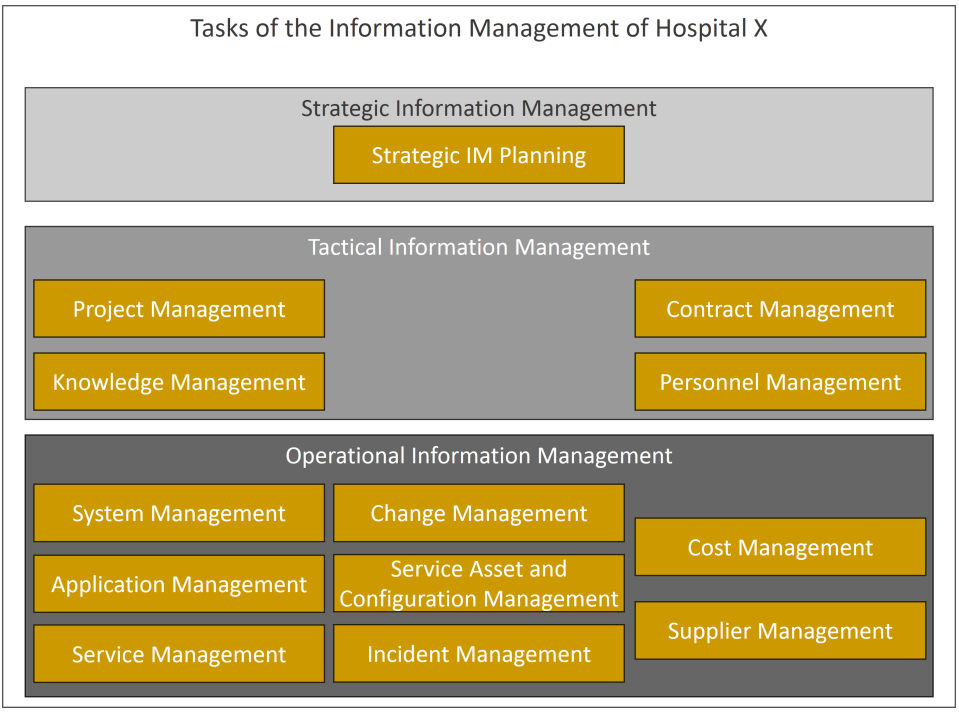


Fig. 6: The identified tasks associated with strategic, tactical and operational IM

Within the strategic IM, we identified the top-level task strategic IM planing. The tactic IM comprises the top-level tasks project management, knowledge management, contract management and personal management. The operational IM comprises the top-level tasks system management, application management, service management, change management, service asset and configuration management, incident management, cost management and supplier management.

The evaluation of the analysis is still ongoing. As a first preview, we want to show the results on the process change management.

4.1 Change Management Process

Figure 7 shows the 3LGM^2 perspective of the modeled change management process in hospital X. The model represents the as-is state of the change management process found in hospital X. The change management process consist of 8 tasks, uses 4 entity types and utilize 3 tools. Tasks, that are identical to the ITIL reference model can be recognized by a green color. The color blue is the standard color for entity types.

The functional and logical tool layers are explained in the following subsections. Since the hardware is not of interest, the physical tool layer will not be discussed.

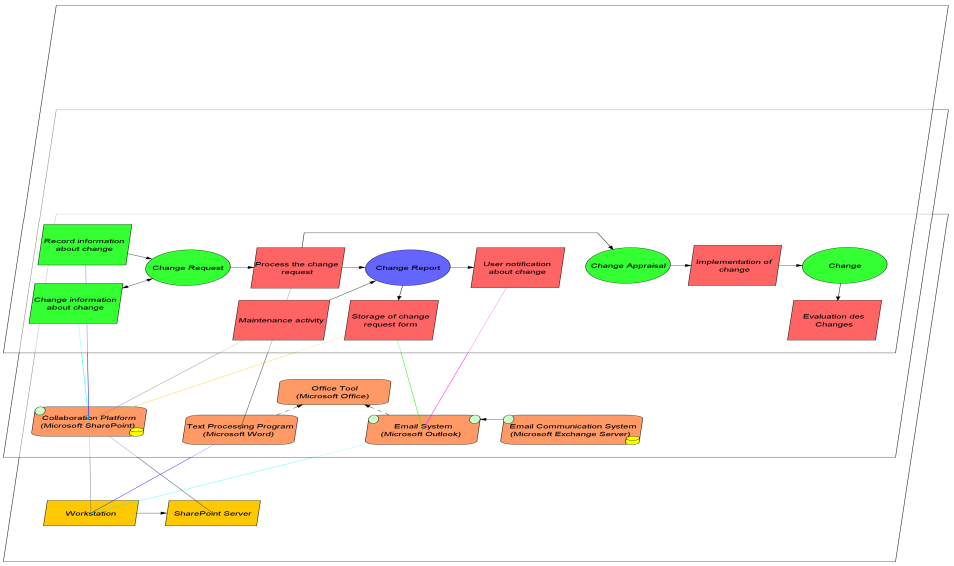


Fig. 7: The identified change management process

4.2 Identified Tasks and Entity Types of Change Management Process

Table 3 shows the identified tasks as well as the consumed and modified entity types of each task. Tasks and entity types correspond to Figure 7.

Whenever a user needs a new or changed functionality, s/he creates a change request using a web based form (T-003 `record information about change`). The output type `change request` is created. In case the change needs a modification, the task T-004 `change information about change` allows editing the `change request`. After the `change request` is received, it is checked for its necessity in task T-005 `process the change request`. When found to be necessary a `change report` with the comment about the change is created. The same entity type `change report` is produced by T-006 `maintenance activity`, which specifies general changes during the monthly maintenance. Then a message about the assessed `change request` is sent to the users in task T-008 `user notification about change`. The `change request` form is now saved for a division-wide access in task T-007 `storage of change request form`. If the `change request` is accepted in the form of a `change appraisal`, the change with the necessary plannings is implemented within the task T-009 `implementation of change`.

As soon as the change is available, its impact can be assessed during the `change evaluation` after its deployment in operations as this is task T-010 `evaluation of change`.

ID	Task	Entity types consumed	Entity types modified
T-003	Record information about change		change request
T-004	Changing information about change	change request	change request
T-005	Process the change request	change request	change report
T-006	Maintenance activity		change report
T-007	Storage of change request form	change report	
T-008	User notification about Change	change report	change appraisal
T-009	Implementation of Change	change appraisal	change
T-010	Evaluation of change	change	

Tab. 3: Entity types used within change management task

4.3 Identified Tools of Change Management Process

The identified tasks of the change management are supported by a collaboration platform (Microsoft Sharepoint), a text editing office system (Microsoft Word) and E-Mail communication (Microsoft Outlook). The mapping between task and tool is summarized in Table 4.

Tool	Task
Microsoft Sharepoint	Record information about change (T-003), Changing information about change (T-004), Maintenance activity (T-006), Storage of change request form (T-007)
Microsoft Word	Process the change request (T-005)
Microsoft Outlook	Storage of change request form (T-007), User notification about change (T-008)

Tab. 4: Tools supporting tasks within change management process

It becomes obvious, that not all of identified tasks performed within the change management are supported by tools. E.g. the evaluation of a change is not supported by a tool. With regard to traceability, a documentation of the change evaluation should be considered.

4.4 Relation to the ITIL Reference Model

The performed system analysis has identified three tasks that have a direct relation to ITIL processes. The interviewees are the source of information about the processes relations to ITIL, i.e. which ITIL elements are used in the process. Obviously the quality of such statements depends on the interviewees' ITIL knowledge. Two of the interviewed persons have an ITIL expert certificate, all others have an ITIL foundation certificate.

Tasks that have a relation to ITIL within the analyzed hospital, are the change management process, the service asset and configuration management and incident management. All other tasks of the ITIL reference model are not implemented in the hospital.

Tasks and entity types in Figure 7, that are used in the hospital, but are not available in the reference model, are marked as red in Figure 7.

As Figure 7 shows for the change management process that the tasks `Record information about change` and `Change information about Change` are identical to the ITIL reference model in Figure 4. Identical entity types are `change request`, `change appraisal` and the `change` itself.

The ITIL reference model in Figure 4 however does not show e.g. the task `Implementation of change`. This task however, is covered by the ITIL standard so that the reference model must be extended. In principle there might be tasks in the hospitals IM, that do not have any relationship to ITIL. These tasks must not be subject to any reference model extension.

Based on this modeling an assessment of the model can be done. The reference model by Issler covers the ITIL v3 2007. However, there are changes to the current version ITIL 2011. This is subject for further investigation.

There are two aspects of model comparability. First, the hospital model to the reference model and second, two or more hospital models with each other.

In this case the reference model and the hospital's model were easy to compare. In general the model comparison is more difficult. For example, if the same task in the reference model and hospital model in their input and output types.

5 Conclusion

This paper presents the application of the system analysis method according to Ammenwerth et al. to analyze and assess the tasks, entity types and tools used within the IM in a hospital.

The results of the system analysis are 1) a description of identified tasks and entity types and tools and 2) a 3LGM^2 model. The model shows what tools contribute to which IM tasks, how these tasks relate to ITIL and what kind of entity types are used.

As one example, the paper describes the task change management in greater detail.

Based on our experiences the combination of a task-oriented approach with interviews as elicitation method and 3LGM^2 as modeling technique is helpful and can thus also be used by other organizations.

In future work we will finish the evaluation of the analysis and then apply the methods to other hospitals. From this we will derive differences and similarities in the IM of the hospitals and to ITIL. We will use the models to evaluate the integration of tools in IM, especially to identify tasks of strategic, tactical and operational IM with missing or inappropriate tool support. Then we start in-depth interviews to elicit requirements for an appropriate IT support for the CIO.

6 Acknowledgment

This work was supported by the DFG (German Research Foundation) under the Project SNIK: Semantic Network of Information Management in Hospitals, Grant no. 1605/7-1 and 1387/8-1.

References

[Am15] Ammenwerth, Elske; Haux, Reinhold; Knaup-Gregori, Petra; Winter, Alfred: IT-Projektmanagement im Gesundheitswesen. Schattauer, Stuttgart, 2., voll. überarb. und erw. aufl. edition, 2015.

[Ar13] Arraj, Valerie: ITIL®: the basics. The APM Group and The Stationary Office, 2013. In: https://www.axelos.com/CMSPages/GetFile.aspx?guid=6117a138-b873-43b9-9fea-372bf6b26bfe. Date visited 18. Dec. 2014.

[Ho11] Hoerbst, Alexander; Hackl, Werner O; Blomer, Roland; Ammenwerth, Elske: The status of IT service management in health care - ITIL® in selected European countries. BMC Medical Informatics and Decision Making, p. 76, 2011.

[Is08] Issler, Lutz: Informationsverarbeitung im Informationsmanagement: Modellierung von Aufgaben und verarbeiteten Informationen am Beispiel der Information Technology Infrastructure Library (ITIL). Dissertation, Medizinische Fakultät der Universität Leipzig, 2008.

[Ja09] Jahn, Franziska; Ißler, Lutz; Winter, Alfred; Takabayashi, Katsuhiko: Comparing a Japanese and a German Hospital Information System - Assessing Structural Quality by 3LGM2 Models. Methods of Information in Medicine, pp. 531–539, 2009.

[PK04] Paech, Barbara; Kohler, Kirstin: Task-driven requirements in object-oriented development. In (J. D. Leite, J, ed.): Doorn, Jorge H. Perspectives on Software Requirements., volume 753, chapter 3 - TASK-D, pp. 1–25. Kluwer Academic, 2004. Print., Boston, MA, 2004.

[Ra11] Rance, Stuart: ITIL® Service Transition. The Stationary Office, London (UK), 2011.

[St13] Steinberg, Randy: ITIL® Service Operation. The Stationary Office, London (UK), 2013.

[WBW03] Winter, Alfred; Brigl, Birgit; Wendt, Thomas: Modeling Hospital Information Systems (Part 1): The Revised Three-layer Graph-based Meta Model 3LGM2. Methods of Information in Medicine, pp. 544–551, 2003.

[We03] Wendt, Thomas; Häber, Anke; Birgit, Brigl; Winter, Alfred: Modeling Hospital Information Systems (Part 2): Using the 3LGM2 Tool for Modeling Patient Record Management. Methods of Information in Medicine, pp. 256–267, 2003.

[Wi07] Winter, Alfred, Brigl Birgit; Funkat, Gert; Häber, Anke abd Heller, Oliver; Wendt, Thomas: 3LGM2-Modeling to support management of health information systems. Int J on Med Inform, pp. 145–150, 2007.

[Wi11] Winter, Alfred: Health information systems. Springer, London [u.a.], 2. ed. edition, c2011.

[WWB15] Wulf, Jochen; Winkler, Till J.; Brenner, Walter: Measuring IT Service Management Capability: Scale Development and Empirical Validation. 2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Business-IT-Alignment in Gemeinden – Qualitative Forschung anhand dreier größerer Berner Gemeinden

Konrad Walser, Philipp Christen, Sabrina Maag, Andrin Regli[1]

Abstract: Durch die elektronische Vernetzung und die zunehmende Komplexität der IT in Verwaltungsorganisationen und insbesondere Gemeinden nimmt die Bedeutung des Business-IT-Alignments (BIA) zu. Je besser die Ausrichtung der IT am Business, so die Annahme, desto höher ist die Nutzenstiftung durch die IT. Allerdings wird dem BIA sowohl von Firmen als auch öffentlichen Verwaltungen häufig immer noch zu wenig Beachtung geschenkt [LB99], [LPB99]). Eine frühere Untersuchung zeigte, dass bei großen Gemeinden das BIA in der Regel besser ausgebildet ist als bei kleinen [BET13], [WEB14]. Aus diesem Grund beschäftigt sich der vorliegende Beitrag mit dem Thema BIA bei drei großen anonymisierten Berner Gemeinden A, B und C. Deren BIA wird qualitativ und auf Basis eines auf Basis der Literatur abgeleiteten spezifischen Reifegradmodells für das BIA in Gemeinden untersucht. Basierend auf Letzterem erfolgt die Ableitung eines Interviewleitfadens. Das Ziel des vorliegenden Beitrags ist es, die Ausprägungs- und Wirkungsarten des strategischen und operativen BIA in großen Gemeinden vertiefter zu untersuchen. Im Hauptteil des vorliegenden Beitrags werden die verschiedenen BIA-Ausprägungen der untersuchten Gemeinden analysiert. Mittels des entwickelten Reifegradmodells ist feststellbar, welche Ausprägung das BIA den Gemeinden hat. Aus der Analyse geht hervor, dass die Reifegradkriterien bezogen auf die Informatik-Abteilungen der untersuchten Gemeinden einen tiefen bis mittleren Reifegrad aufweisen. Dies ist u.a. darauf zurückzuführen, dass die vorhandenen Strategie-Dokumente keinen Bezug zu Legislaturzielen der Gemeinden haben und dazu auch keinen Beitrag leisten. Weiter mangelt es teilweise an der Zusammenarbeit und der Kommunikation mit den internen Informatikkunden (Direktionen). Auch operative Gremien werden als Mittel zum BIA nicht bestmöglich eingesetzt. Zudem sehen sich die Informatik-Abteilungen häufig selbst nur als Dienstleister und Enabler, was ihr Verhalten beeinflusst und sie im proaktiven Handeln hemmt.

Keywords: Business-IT-Alignment, Gemeinden, E-Government, IT-Komplexität, Fallstudien, Reifegradmodell, Reifegradmessung, qualitative Forschung.

1 Einleitung

1.1 Problemstellung

Das BIA stellt in der Wirtschaftsinformatik kein neues Thema dar. Etwa wurde es 1993 von [HV93] als Strategic Alignment eher als theoretisches Artefakt konzipiert. Überdies erfolgten basierend auf den konzeptionellen Überlegungen von [HV93] sowie weiteren Forschern ([AJP04], empirische Ergebnisse [CKA05], [SC01] zusätzliche Forschungen

[1] Berner Fachhochschule, E-Government-Institut, Brückenstrasse 73, CH-3005 Bern, konrad.walser@bfh.ch, philipp.christen@students.bfh.ch, sabi.maag@gmail.com, andrin.regli@students.bfh.ch.

und Entwicklungen von BIA-Ansätzen beispielsweise auf Basis von IT-Governance-Rahmenwerken [DG09]) weitere Entwicklungen, die jedoch nicht zu einer einheitlichen Vorstellung zum BIA geschweige denn zu einer einheitlichen Definitionen führten. Unklar ist ferner auch welche Instrumente, Artefakte, Werkzeuge und Komponenten dazu typischerweise zum Einsatz gelangen. Am ehesten ist dies noch der Fall beim Strategic Alignment welches den Abgleich von Unternehmens- und IT-Strategie zum Inhalt hat. Eine zentrale Intention des BIA kann daher sein, dass der Abgleich der IT mit dem Geschäft zu einem höheren Nutzen der IT für das Geschäft und damit letztlich zu mehr Erfolg der Unternehmung am Markt führt. Was ist ein entsprechendes Ziel des BIA in der öffentlichen Verwaltung? Ist dies mehr Effizienz und/oder mehr Effektivität im Verwaltungshandeln zu erreichen? Diese Frage ist längerfristig zu klären und nicht Gegenstand der Untersuchung in diesem Beitrag. Hypothetisch gesehen dürfte beides der Fall sein. Ausgehend von der Entwicklung der Technologie, aber auch angesichts der zunehmenden E-Government-Verbreitung unter den Gemeinden und damit der Vernetzung, steigt die IT-Komplexität generell oder etwa die IT-Sourcing-Komplexität. Damit kommt dem BIA in der öffentlichen Verwaltung eine immer höhere Bedeutung zu, was in erstaunlichem Gegensatz zu wissenschaftlichen Erkenntnissen steht ([BET13], [WEB14]). Die Bedeutung der Zusammenarbeit in Netzwerken zwischen den Gemeinden als autonome Beteiligte der öffentlichen Verwaltung nimmt zu. Die Komplexität des IT-Managements in Kooperationen wirkt sich auch auf das BIA aus. Die Bedeutung des BIA ist gemäß Untersuchungen insbesondere aus der Sicht von kleinen und mittleren Gemeinden heute klein (vgl. [BET13], [CMR14], [WEB14]). In größeren Gemeinden ist die Bedeutung des BIA größer, daher wurde die vorliegende Studie bei ebensolchen durchgeführt [CMR14]. Aus theoretischer Sicht fehlt heute wie erwähnt eine verlässliche Grundlage, um die Bausteine eines BIA auf strategischer, taktischer und operativer Ebene auszumachen, konsistent zueinander einzusetzen und zu steuern. Ferner fehlt in Gemeinden vielfach das Bewusstsein dafür, dass IT heute mehr sein kann als ein reines Unterstützungsinstrument, nämlich ein Enabler für ein zunehmend innovativeres Verwaltungsgeschäft. Maßgeblich zu dieser Erkenntnis beitragen kann ein BIA. Die vorliegende Studie analysiert Wirkungsmechanismen und -zusammenhänge eines Business-IT-Alignements auf Basis qualitativer Forschung. Neben der Forschung von [BET13], [WEB14], [CMR14] ist hier weitere Forschung erforderlich. All diese Überlegungen führen zu folgenden Forschungsfragen: Wie findet das BIA in großen Gemeinden statt? Wie ist das BIA bei großen Berner Gemeinden ausgeprägt? Warum werden keine, respektive nur wenige IT-Standards verwendet [BET13]?

1.2 Zielsetzung

Die Zielsetzungen des vorliegenden Beitrags lauten: Aufsetzend auf einer vorangehenden Studie quantitativer Art durch die Autoren [BET13], [WEB14] erfolgt in einem ersten Schritt eine Ergänzung der entsprechenden quantitativen Studie um qualitative Leitfadeninterviews und eine Untersuchung zu Ausprägungs- und Wirkungsarten des BIA in (großen) Gemeinden. In einem zweiten Schritt werden die verschiedenen BIA-Ausprägungen der untersuchten Gemeinden miteinander verglichen.

Mittels eines eigens entwickelten Reifegradmodells wird entsprechend festgestellt, welchen BIA-Reifegrad die untersuchten Gemeinden haben.

2 Theoretische Grundlagen

Der Begriff Business-IT-Alignment (BIA) wird in der Literatur uneinheitlich definiert und synonym mit Begriffen wie folgt verwendet: IT Alignment, Strategic Alignment. Letzterer eignet sich für die weitere Verwendung nicht, da er nur die Abstimmung von Geschäfts- und IT-Strategie und nicht ein ganzheitliches Zusammenwirken von Business und IT ([BEH10], S. 36) intendiert. [Ni04] versteht unter dem Begriff Business-IT-Alignment das Folgende: „In general, IT-business alignment […] concerns the degree of congruence of an organization's IT strategy and IT infrastructure with the organization's strategic business objectives and infrastructure." ([Ni04], S. 80). Mit BIA ist also der Grad an Kongruenz zwischen der IT-Strategie und -Infrastruktur und der Geschäftsstrategie und -infrastruktur gemeint. [Lu00] spricht bei BIA nicht von Kongruenz, sondern von Harmonie zwischen der Strategie, Zielen und Bedürfnissen von IT- und Geschäftsseite: „Business-IT alignment refers to applying Information Technology (IT) in an appropriate and timely way, in harmony with business strategies, goals and needs." ([Lu00], S. 3). Auch [BEH10] definieren BIA in ähnlicher Weise, teilweise operativer, als: „[…] die wechselseitige Abstimmung von Zielen, Strategien, Architekturen, Leistungen, Prozessen und der Unternehmenskultur zwischen IT-Bereich und Fachbereich in Unternehmen" ([BEH10], S. 36). Diese letzte Definition zeigt, dass bei BIA nicht nur die Strategie von Bedeutung ist, sondern auch weitere Faktoren wie Kultur, Leistung, Architektur, Prozesse, usw., die ins BIA miteinbezogen werden müssen.



Abbildung 1. Unterteilung BIA nach [GKB07], zitiert nach [BEH10], S. 42.

Dies führt bezüglich einer aktiven Gestaltung des BIA zu erhöhter Komplexität und zur bis heute nicht schlüssig beantworteten Frage, wodurch und durch welche Werkzeuge dies erfolgen soll. Die Thematisierung von BIA – insbesondere des strategischen

Alignments – geht im Wesentlichen und wie erwähnt auf [HV93] zurück (S. 476). Deren entwickeltes Modell wird auch als Strategic Alignment Model (SAM) bezeichnet. Dabei liegt die Verlinkung der Business- mit der IT-Strategie als externe sowie der Business-Organisation mit der IT-Organisation als interne Übereinstimmung zugrunde. Für die Umsetzung der jeweiligen Strategien müssen die beiden Levels (extern/intern) jeweils aufeinander abgestimmt werden. Dies bedeutet, dass beispielsweise die für die Umsetzung der IT-Strategie relevanten Architekturen, Prozesse und Fähigkeiten der Mitarbeitenden zur Verfügung stehen oder aufgebaut werden. Eine Untersuchung zum BIA in der öffentlichen Verwaltung hat insbesondere deren spezifischen Kontext zu berücksichtigen. Öffentliche Verwaltungen haben in der Regel keine (Unternehmens-)-Strategie, wie dies bei Unternehmen der Fall ist, sondern Legislaturziele oder -richtlinien. In Letzteren wird die (politische oder strategische) Stoßrichtung der Gemeinde und werden entsprechende Ziele definiert. Unter dem Begriff Legislatur wird der Zeitraum verstanden, während dem die gewählten Volksvertreter (in Exekutive und Parlament) ihr Amt innehaben, in der Schweiz in der Regel vier Jahre. Danach erfolgen Wahlen zur nächsten Legislatur [Vi11]. Unter Berücksichtigung dieses Sachverhalts wird für die vorliegende Arbeit die folgende Definition des Begriffs BIA verwendet: BIA bei Gemeinden ist die Ableitung der mittel- bis langfristigen IT-Stoßrichtung aus den Legislaturzielen und somit die gesamtheitliche Ausrichtung der IT (Mitarbeitende, Prozesse, Architektur, Ziele, usw.) auf die IT-Bedürfnisse der Gemeinden mit dem Ziel, einen Geschäftsmehrwert im Sinne einer Effizienz- und/oder Effektivitätssteigerung in der (vernetzten) Verwaltung zu schaffen. Das BIA lässt sich in verschiedene Teildisziplinen unterteilen, u.a. in operatives und strategisches Alignment, wie dies in der folgenden Abbildung dargestellt wird ([BEH10], S. 37). Gemäß [WBS14] gibt es neben dem strategischen und dem operativen BIA auch noch eine taktische Abstimmung zwischen den Geschäftsbereichen und der Informatik. Sie betrifft Portfolios und die Infrastruktur, während das strategische BIA Ziele, Strategien und Pläne beinhaltet und sich das operative BIA auf das Tagesgeschäft, Projekte und Geschäftsprozesse konzentriert. Auf allen drei Ebenen kann zusätzlich zwischen strukturellem Alignment, wie Organisation, Strukturen und Abläufe, sowie sozialem Alignment, das u.a. die Kommunikationsqualität, gegenseitiges Vertrauen und bereichsübergreifendes Wissen beinhaltet, unterschieden werden [WBS14]. Strategisches BIA beschreibt die Abstimmung der IT- und der Geschäfts-Strategie. Damit ist nicht zwingend eine einseitige Ausrichtung der IT-Strategie auf die Geschäftsstrategie gemeint. Die Geschäftsstrategie kann durchaus auch durch die IT inspiriert werden. Damit ein solches Zusammenspiel funktioniert, ist es jedoch wichtig, dass die IT die Funktionsweise des Business versteht und eine partnerschaftliche Beziehung zwischen dem Business und der IT besteht. Außerdem ist es vorteilhaft, wenn das Senior-Management die Wichtigkeit der IT versteht und auch entsprechend unterstützt ([Lu00], S. 6ff.). Im Gegensatz zum strategischen BIA fokussiert das operative BIA auf die wechselseitige Koordination und Interaktion der Fach- und IT-Abteilungen und deren Mitarbeitenden auf operativer Ebene. Die IT richtet sich dabei an den mittelfristigen Zielen der Fachabteilungen aus. Die Mitarbeitenden im operativen Geschäft sind mit dieser Art von BIA täglich konfrontiert und schaffen durch ihre Arbeit einen Wertbeitrag für das Unternehmen ([HD12], S.

2). Wie aus der obigen Abbildung hervorgeht, lässt sich das operative BIA in drei Dimensionen unterteilen: Kognitive Dimension, Interaktionsdimension sowie Qualitative Dimension. In der kognitiven Dimension sind die Faktoren Respekt, Vertrauen und gemeinsame Ziele vorhanden, während in der Interaktionsdimension die Kommunikation und die Art der Zusammenarbeit beschrieben werden. Die qualitative Dimension beschreibt vorwiegend wissensbasierte Faktoren, beispielsweise inwiefern und wie tief Kenntnisse und Verständnis über Prozesse und Vorgänge der anderen Einheit vorhanden sind ([GKB07], S. 31ff.). Gemäß [BWF07] wirken alle drei Dimensionen auf den Geschäftswert (Business Value) ein und können einen positiven Einfluss auf die Prozess-Performance leisten. Jedoch wird festgestellt, dass die qualitative Dimension wichtiger zu sein scheint. Insbesondere das Wissen der IT über die Business-Prozesse kristallisiert sich im operativen Business-IT-Alignement als wichtigster Faktor heraus ([BWF07], S. 1-8). Wie in den obigen Abschnitten bereits angedeutet wurde, braucht es für ein erfolgreiches BIA auf operativer Ebene ein gutes Projekt-Portfoliomanagement. Dieses sorgt für die Umsetzung der aus der Unternehmensstrategie abgeleiteten IT-Strategie und steuert die Veränderungen innerhalb der gesamten Unternehmung. Durch die Priorisierung von Projekten anhand ihres Business Cases, dem strategischen Wertbeitrag und ihrer Risiken kann das Projekt-Portfoliomanagement das BIA umsetzen ([HD12], S. 3). Wie die folgende Abbildung zeigt, werden zur Umsetzung des BIA auf der operativen Ebene Werkzeuge und Prozesse benötigt. Die wichtigsten Werkzeuge dazu sind u.a. der Servicekatalog, die mit den Fachabteilungen vereinbarten Service Level Agreements (SLA), regelmäßige Service Level Reports und Projekte, welche die Anforderungen des Business umsetzen. Das Service Level Management und das Service Catalogue Management, welche das BIA auf operativer Ebene sicherstellen, sind Teil des IT-Frameworks ITIL 2011 edition. Auf dieses wird in einem Kapitel weiter unten eingegangen. Wenn in einem Unternehmen ein BIA stattfindet und die IT die Geschäftsseite optimal unterstützen kann, resultieren daraus verschiedene Vorteile.

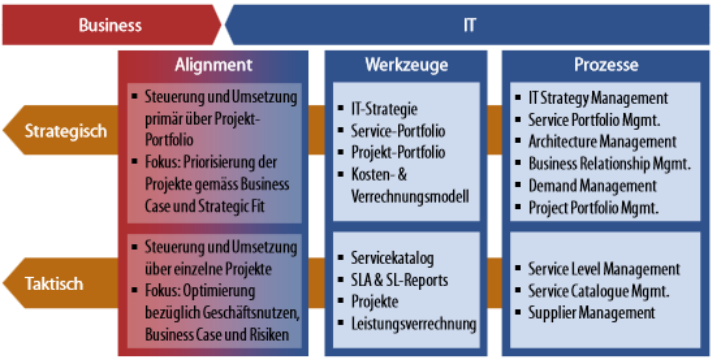


Abbildung 2: Steuerung des taktischen BIA durch Werkzeuge und Prozesse ([HD12], S. 3) mehrheitlich basierend auf ITIL 2011 edition.

Gemäß ([Be06], S. 3) erzielen Organisationen (hier sind Unternehmen im Markt gemeint) durch ein optimales BIA die folgenden Nutzen: Kostenreduktion, Standardisierte

Prozesse, Produktivitätssteigerung, Workflow- und Kommunikations-Verbesserungen, Einhaltung von Service Levels, verbesserte Kontrollmechanismen für Risiken, Implementierung von neuen Business Strategien, Ermöglichung von organischem und Aquisitions-Wachstum sowie Wettbewerbsvorteile durch neue Technologien. Weiter kann BIA transparente und effektive Strukturen herbeiführen. Dies hat neben den bereits erwähnten Vorteilen auch positive Konsequenzen für die Allokation von Ressourcen. Zudem wird eine stabilere Planungsbasis für die Zukunft geschaffen und es werden interne/externe Anwender zufriedengestellt.

3 Entwicklung eines Reifegradmodells für die BIA-Maturitätsmessung bei Gemeinden

Für die in diesem Beitrag erfolgende Bewertung des Reifegrades des BIA der Berner Gemeinden wird ein eigenes fünfstufiges Reifegradmodell erstellt. Das für die vorliegende Arbeit eigens entwickelte Reifegradmodell lehnt sich stark an die fünf Stufen des CMMI-Reifegradmodells (vgl. [CM11]) an. Damit der Reifegrad des BIA beurteilt werden kann, müssen zuerst geeignete Kriterien definiert werden. Diese lassen sich aus den zu untersuchenden drei wichtigsten Themengebieten, die sich aus den Theoretischen Grundlagen ergeben, ableiten: allgemeines BIA, strategisches BIA sowie operatives BIA. Nicht als Themengebiet in das Reifegradmodell aufgenommen werden IT-Standards und -Frameworks, da diese gemäß ([BET13], S. 37) bei den Gemeinden kaum eingesetzt werden und sich folglich kein Reifegrad daraus ableiten lassen würde. Pro Themengebiet werden drei Reifegradkriterien für jeweils fünf Stufen definiert. Die Stufen entsprechen den Leveln von CMMI und werden diesen wie folgt zugeordnet (ergänzt in Klammern um CMMI-Charakterisierungen der Stufen gemäß ([CMoJ]): 1. sehr tief = initial (Prozess unvorhersehbar, schlecht gesteuert und reakiv); 2. tief = managed (Prozess charakterisiert für Projekte, wird gemanaged); 3. mittel = defined (Prozess charakterisiert für Organisation, proaktiv); 4. hoch = quantitatively managed (Prozess quantitativ gesteuert und gemessen); 5. sehr hoch = optimizing (Fokus auf kontinuierliche Prozessverbesserung). Die für das BIA von Gemeinden definierten Reifegradkriterien und -stufen sind in den nachfolgenden Tabellen aufgeführt.[2] Die erste Tabelle bezeichnet die Reifegradekriterien und -charakterisierungen für das allgemeine Alignment.

[2] Es ist zu beachten, dass die Begriffe IT und Informatik sowie Abteilungen und Direktionen synonym verwendet werden.

Kriterium		Reifegradstufen				
		sehr tief	**tief**	**mittel**	**hoch**	**sehr hoch**
1.1	Von der Informatik angebotene Dienstleistungen	Die Leistungen der Informatik sind nicht als Dienstleistungen aufgebaut.	IT- Dienstleistun gen werden angeboten, sind aber nicht ein- heitlich.	Die angebotenen IT-Dienstleistungen sind Gemeindeweit einheitlich	Gewinn und Umsatz der angebotenen IT-Dienstleistungen sind messbar.	Die angebotenen IT-Dienstleistungen werden laufend verbessert.
1.2	Einkauf externer IT-Dienstleistungegen	Der Einkauf von IT-Leistungen erfolgt dezentral, ungesteuert und ad-hoc.	Der Einkauf von IT- Dienstleistungen erfolgt zentral, ist aber nicht standardisiert.	Der Einkauf von IT- Dienstleistungen ist Gemeinde-weit einheitlich und transparent.	Die Investitionen in externe IT- Dienstleistungen sind messbar.	Der Einkauf von externen Dienstleistungen wird laufend verbessert.
1.3	Wahrnehmung von IT-Leistungsbezügern als Kunden	Der Begriff Kunde existiert bei der Informatik nicht.	Die Leistungsbezüger werden nicht gleich behandelt. Es gibt keine einheitliche Kunden-	Einheitliche Kundendefinition vorhanden, welche Leistungsbezüger in Kategorien eint.	Die Kunden sind kategorisiert und es besteht ein aktives Kundenmanagement.	Die Kundenbeziehungen werden laufend verbessert.

Abbildung 3: Generisches BIA und entsprechende Maturitätsstufen-Charakterisierung.

Die folgende Tabelle bezeichnet die Reifegradekriterien und -charakterisierungen für das strategische Alignment.

Kriterium		Reifegradstufen				
		sehr tief	**tief**	**mittel**	**hoch**	**sehr hoch**
2.1	IT-Strategie	Es ist keine IT- Strategie vorhanden.	Eine IT-Strategie ist partiell vorhanden aber nicht konsistent.	Eine IT-Strategie ist in einem Dokument vorhanden.	Die Umsetzung der IT-Strategie ist messbar.	Die IT-Strategie und deren Umsetzung werden laufend verbessert.
2.2	Bezug und Beitrag der IT-Strategie zu den Legislaturzielen	Die IT-Strategie hat keinen Bezug zu den Legislaturzielen.	Die IT-Strategie hat nur einen partiellen oder indirekten und keinen konsistenten Bezug zu den Legislaturzielen.	Die IT-Strategie hat einen einheitlichen und transparenten Bezug zu den Legislaturzielen.	Der Bezug und Beitrag der IT-Strategie zu den Legislaturzielen ist messbar.	Der Bezug und Beitrag zu den Legislaturzielen wird laufend optimiert.
2.3	Rolle der Informatik	Die Informatik nimmt keine (bedeutsame) Rolle ein.	Die Rolle der Informatik ist nur teilweise definiert und bekannt.	Die Rolle der Informatik ist definiert und Gemeinde-weit bekannt.	Die Zielerreichung der IT gemäß ihrer Rolle wird regelmäßig gemessen.	Die Rolle der Informatik wird überprüft und wo sinnvoll angepasst.

Abbildung 4: Strategisches BIA und entsprechende Maturitätsstufen-Charakterisierung.

Die folgende Tabelle bezeichnet die Reifegradekriterien und -charakterisierungen für das operative Alignment.

Kriterium		Reifegradstufen				
		sehr tief	**tief**	**mittel**	**hoch**	**sehr hoch**
3.1	Zusammenarbeit mit den Fachabteilungen im Alltag und in Projekten	Es findet keine Zusammenarbeit mit den Abteilungen statt.	Es findet Zusammenarbeit statt, sie ist jedoch nicht einheitlich und meist ad-hoc.	Die Zusammenarbeit ist Ge- meindeweit standardisiert und transparent.	Die Zusammenarbeit mit den Abteilungen wird aktiv gefördert.	Die Zusammenarbeit mit den Abteilungen wird laufend optimiert.
3.2	Austausch in operativen Gremien	Es existieren keine gemeinsamen operativen Gremien.	Es existieren Gremien, mit denen ein parti- eller Austausch stattfindet. Diese sind meist ad-hoc organisiert.	Der regelmäßige Austausch in den Gremien ist Gemeinde-weit standardisiert und transparent.	Der regelmässige Austausch in gemeinsamen Gremien wird aktiv gefördert.	Der regelmäßige Austausch in Gremien und die Gremien selbst werden laufend optimiert.
3.3	Kommunikation mit den Fachabteilungen	Es besteht keine Kommunikation zwischen der Informatik und den Abteilun- gen.	Die Kommunikation mit den Abteilungen ist nicht einheitlich und transparent.	Die Kommunikation mit den Abteilungen ist Gemeinde-weit standardisiert und transparent.	Die Kommunikation mit den Abteilungen wird aktiv gefördert.	Die Kommunikation mit den Abteilungen wird laufend optimiert.

Abbildung 5: Operatives BIA und entsprechende Maturitätsstufen-Charakterisierung.

Eine tabellarische Beschreibung und ein Verglich der drei untersuchten Gemeinden ist in der folgenden Abbildung dargestellt.

	Gemeinde A	**Gemeinde B**	**Gemeinde C**
Anzahl Einwohner[3]	54‘562	15‘917	14‘130
Konstitution organisatorisch	Stadtrat und Gemeinderat, Führung durch Stadtpräsident	Gemeinderat, Stadtrat und Stadtpräsidenten	Gemeinderat
Art der Verwaltungs-Direktionen	Präsidialdirektion, Finanzdirektion, Direktion Soziales und Sicherheit, Direktion Bildung, Kultur und Sport sowie die Di-rektion Bau, Energie und Umwelt	Baudirektion, Bildungsdirektion, Einwohner- und Sicherheitsdirektion, Finanzdirektion, Präsidialdirektion und die Sozialdirektion	Präsidiales, Finanzen, Bau und Planung, Bildung und Kultur, Sicherheit und Liegenschaften sowie Soziales und Jugend
Positionierung Informatik in Organisation	Informatik und Logistik der Stadt A ist als eigenständige Abteilung innerhalb der Finanzdirektion	Informatik zur Finanzdirektion gehörig	Informatik zur Finanzdirektion gehörig

[3] Stand Ende 2013.

	Gemeinde A	Gemeinde B	Gemeinde C
Zahl der IT-Mitarbeiter	36 FTEs, davon 27 in der Informatik	3 FTEs	0; Je Abteilung gibt es eine zu-ständige Person für die Informatik. IT-Outsourcing an Talus Informatik AG, Seedorf.
IT-Budget pro Jahr	7.5 Mio. CHF	850‘000 bis 900‘000 CHF; jährliche Informatikkosten betragen pro Arbeitsplatz ca. 5‘000 CHF	590‘000 CHF
Zahl betreuter Arbeitsplätze	800, in gewissen Bereichen auch Unterstützung kleinerer Gemeinden	250, inkl. Arbeitsplätze Gemeinde X	100 User, sofern dies als Äquivalent für Arbeitsplätze bezeichnet werden kann.
Zahl der Mitarbeiter der Gemeinde	2‘500 zu 1‘400 FTEs		
Bemerkungen		2004 Umstellung auf Citrix, Virtualisierung; pro Einwohner der Stadt B fallen pro Jahr Informatikkosten in Höhe von 35 - 80 CHF an.	Einsatz des New Public Management Konzepts in der Gemeindeverwaltung; Kosten pro Arbeitsplatz: CHF 6‘600, davor durchschnittlich zwischen 7‘200 und 5‘600 CHF.

Abbildung 6: Charakterisierung der drei untersuchten Berner Gemeinden.

4 Methodisches Vorgehen

Die Untersuchung erfolgte als qualitatives Vorgehen mit Leitfadeninterviews, basierend auf dem bereits vorhandenem Wissen durch eine Vorgängeruntersuchung von [BET13] sowie [WEB14] und dem konkretisierten Reifegradmodell für die öffentliche Verwaltung. Der Fokus der Untersuchung lautete wie folgt: Ausprägungs- und Wirkungsarten des BIA in Gemeinden. Fünf Forschungsphasen werden für das Forschungsvorgehen wie folgt unterschieden: Problemerfassung (Literaturrecherche, Konkretisierung der Forschungsfragen), Informationsgewinnung und Erkenntnisgenerierung (Definition der Stichprobe in Form einer Fallauswahl, Entwicklung eines eigenen BIA-Reifegradmodells zur Operationalisierung der Forschungsfrage, Ableitung eines Interview-Leitfadens aus dem Reifegradmodell, Generierung von Informationen mittels Leitfaden aus Interviews), Resultatvalidierung (Gegenüberstellung der Resultate, Vergleich und Auswertung), Berichtsabfassung. Im vorliegenden Beitrag erfolgt ein Mehrfallstudien-Design (vgl. [Yi09]) mit drei Fallstudien. Je Gemeinde wird das Verhältnis zwischen IT und Geschäft der Gemeinde untersucht. Die Fallauswahl erfolgt nach den folgenden Kriterien: Gemeinden des Kantons Bern, geografische Zugehörigkeit (Stadt/Agglomeration), Größe der Gemeinde sowie Informationsverfügbarkeit (frei zugängliche Daten auf der Gemeindewebseite) sowie Interviewbereitschaft im Hinblick auf die Datenerhebung. Die Webseite des Bundesamts für Statistik stellt eine Liste der Gemeinden der Schweiz zur Verfügung. Mittels Filtern lassen sich die für die Studie möglichen Gemeinden eruieren. Ausgehend vom Kriterium >10‘000 Einwohner resul-

tieren schließlich die folgenden vierzehn Gemeinden: Belp, Bern, Biel/Bienne, Burgdorf, Ittigen, Köniz, Langenthal, Lyss, Münsingen, Muri bei Bern, Ostermundigen, Spiez, Steffisburg, Thun und Worb. Unter Anwendung der übrigen Kriterien resultieren für die Untersuchung nach zwei elektronischen Anschriften mit positiven und negativen Antworten schließlich die zu untersuchenden Gemeinden wie folgt: Gemeinde A, Gemeinde B, Gemeinde C. Die Fragen für den Interviewleitfaden lassen sich teilweise aufgrund der erzielten Resultate der qualitativen Studie von [BET13] zum Thema BIA bei Gemeinden ableiten. Weitere Fragenstellungen ergeben sich aus den erarbeiteten theoretischen Grundlagen, dem selbst entwickelten Reifegradmodell sowie aus den Forschungsfragen. Der Interviewleitfaden kann im vorliegenden Beitrag aus Platzgründen nicht präsentiert werden, ist aber bei den Autoren per Email abrufbar. Aus zeitlichen Gründen war kein Pretest möglich, der Fragebogen musste somit falls nötig im ordentlichen Ablauf der Forschung angepasst werden. Eine Codierung und Inhaltsanalyse entfielen aus Ressourcen- und Zeitgründen, jedoch wurden die Interviewtranskripte der Fallstudien den Interviewten zum Gegenlesen zugestellt.

5 Forschungsergebnisse

Aus Gründen der Ausführlichkeit der Forschungsergebnisse können diese aus Platzgründen nicht unmittelbar im Beitrag präsentiert werden Die Forschungsergebnisse werden daher in einem separaten Dokument präsentiert, das auf Dropbox abgelegt ist. Der entsprechende Link dazu findet sich im Anhang dieses Beitrags.

6 Generische Wertung der Untersuchung

Die folgende Tabelle fasst die Untersuchungsresultate zusammen.

Untersuchungs-Bereich	**Zusammenfassung der Untersuchungsresultate**
Allgemeines BIA	Alle drei Gemeinden nehmen ihre IT-Leistungsbezüger als Kunden wahr und teilen diese in eine interne und externe Kategorie ein Anzahl Kunden ist jeder IT-Abteilung bekannt Kriterien zu den angebotenen und eingekauften IT-Dienstleistungen sind je nach Gemeinde verschieden, aber in sich konsistent IT der Gemeinde A mit tiefen Reifegraden, da weder Angebot noch externer Bezug einheitlich IT der Gemeinde B in diesen Kriterien mit hohen Reifegraden, weil ihre angebotenen Dienstleistungen transparent, Beschaffungs-prozesse standardisiert sowie alles finanziell messbar ist Einheitliche Prozesse und transparente Dienst-leistungen sind auch bei der IT der Gemeinde C existierend, deshalb hier mittlere Reifegrade verzeichnend
Strategisches BIA	Für Kriterien IT-Strategie und Bezug zu Legislaturzielen tiefe bis mittlere Reifegrade vergeben Grund: vorhandene Strategiedokumente nicht auf Legislaturziele Gemeinden ausgerichtet und dazu auch keinen Beitrag leistend Gemeinde A: interne IT-Strategie vorliegend, unabhängig von Legislaturzielen erstellt

Unter-suchungs-Bereich	Zusammenfassung der Untersuchungsresultate
	Gemeinde B: Informatik-Konzept mit strategischen Leitsätzen in Anwendung, jedoch keine verabschiedete IT-Strategie Gemeinde C: am besten abschneidend: im Sinne einer IT-Strategie strategische Grundsätze IT einsetzend Rolle der Informatik: alle drei Gemeinden sich der Rolle bewusst seiend und sie auch Gemeinde-weit in gleicher Art und Weise lebend. Daraus ist indes nicht schließbar, dass eingenommenen Rollen auch diejenigen sind, die Gemeinden strategisch am besten unterstützen Abgleich zwischen Strategie und IT-Rolle müsste vertiefter untersucht werden
Opera-tives BIA	Insgesamt tiefer bis mittlerer Reifegrad Beurteilung Kriterien pro Gemeinde in sich stimmig Gemeinde A: Tiefe Reife, da sich Zusammenarbeit mit Direktionen lediglich auf Support beschränkend und diese bei Informatik-Sitzungen auch nicht allesamt vertreten sind; Kommunikation in der Gemeinde A mangelhaft Gemeinde B: Zusammenarbeit und Kommunikation mit Abteilungen transparent, daher mittlere Reife aufweisend. Zu verbessern wäre das operative Gremium Steuerungsausschuss, mit dem ein nur partieller und nicht immer regelmäßiger Austausch stattfindet Gemeinde C: konsistent mit mittlerer Reife. Abteilungen werden gleichermaßen in Zusammenarbeit miteinbezogen und Kommunikation geschieht Gemeinde-weit transparent und einheitlich. Regelmäßiger Austausch mit Abteilungsleitenden.

Abbildung 7: Zusammenfassung der Resultate über die drei Maturitätsmessungen in großen Berner Gemeinden

Speziell ist anzumerken, dass neben den genannten Gründen auch die Organisation der Gemeinden einen Einfluss auf die tiefen bis mittleren Reifegrade hat. So überwiegt die Politik in praktisch allen Bereichen, wodurch die Interessen anders gelagert sind, als dies beispielsweise bei einer privatwirtschaftlichen Unternehmung der Fall ist. In einer solchen hat der CEO ein großes Interesse daran, die Geschäftsprozesse zu optimieren und Kosten mit Hilfe der IT langfristig zu senken oder etwa Umsätze langfristig zu steigern. Im Gegensatz dazu wollen die Gemeindepräsidenten und -räte sowie involvierte politische Parteien eher dort Geld investieren, wo für die Wähler ein direkter Nutzen besteht. Dies ist selten im Bereich IT der Fall. Politiker wollen ausgehend von Ihren Entscheidungen bei den nächsten Wahlen profitieren. Dies bedeutet, dass der Fokus eines Politikers auf maximal vier Jahre bis zur nächsten Wahl ausgelegt ist. Investitionen in die und Thematisierungen der IT sind bei Politikern unbeliebt und wenig Prestige-wirksam.

7 Zusammenfassung, Handlungsempfehlungen und Ausblick

7.1 Zusammenfassung

Zusammenfassend kann angemerkt werden, bewegen sich die BIA-Reifegrade über alle Dimensionen oder Kriterien hinweg auf einem tiefen bis mittleren Niveau. Es wurde bewusst auf eine numerische Bewertung und mathematische Berechnungen der Maturität

verzichtet, da diese zu einem Durchschnitt führen würde, der in der Praxis wenig brauchbar wäre und keinen Mehrwert bringen würde. Vielmehr scheint es bedeutsam dass die Schwachstellen erkannt, benannt und verbessert werden. Die Untersuchungen zum allgemeinen BIA (Kriterien 1.1-1.3) haben ergeben, dass alle drei Gemeinden ihre IT-Leistungsbezüger als Kunden wahrnehmen und diese in eine interne und externe Kategorie einteilen. Auch die Anzahl der Kunden ist jeder IT-Abteilung bekannt. Die Kriterien zu den angebotenen und eingekauften IT-Dienstleistungen sind je nach Gemeinde verschieden, aber in sich konsistent. Etwas anders sehen die Reifegrade beim strategischen BIA aus. Hier wurden für die Kriterien IT-Strategie und Bezug zu den Legislaturzielen tiefe bis mittlere Reifegrade vergeben. Dies ist hauptsächlich darauf zurückzuführen, dass die vorhandenen Strategie-Dokumente nicht auf die Legislaturziele der Gemeinden ausgerichtet sind und dazu auch keinen Beitrag leisten. Alle drei Gemeinden sind sich bezüglich der Rolle der Informatik bewusst und leben sie auch in gleicher Art und Weise. Jedoch kann daraus nicht geschlossen werden, dass die eingenommenen Rollen auch diejenigen sind, die die Gemeinden strategisch am besten unterstützen. Ein Abgleich zwischen Strategie und IT-Rolle ist vertiefter zu untersuchen. Das operative BIA fällt insgesamt mit einem tiefen bis mittleren Reifegrad aus. Auch hier ist die Beurteilung der Kriterien pro Gemeinde in sich stimmig. Neben den genannten Gründen hat aber auch die Organisation der Gemeinen einen Einfluss auf die tiefen bis mittleren Reifegrade. So überwiegt die Politik in praktisch allen Bereichen, wodurch die Interessen anders gelagert sind, als dies beispielsweise bei einer privatwirtschaftlichen Unternehmung der Fall ist. In einer solchen hat der CEO ein großes Interesse daran, die Geschäftsprozesse zu optimieren und ihre Kosten mit Hilfe der IT langfristig zu senken. Im Gegensatz dazu wollen die Gemeindepräsidenten und -räte sowie involvierte politische Parteien eher dort Geld investieren, wo für die Wähler ein direkter Nutzen entsteht und sie dadurch bei den nächsten Wahlen davon profitieren können. Dies bedeutet auch, dass der Fokus eines Politikers nicht langfristig sondern auf maximal vier Jahre bis zur nächsten Wahl ausgelegt ist. Zudem sind Investitionen in die IT innerhalb der Politik eher unbeliebt und wenig Prestige-wirksam.

7.2 Handlungsempfehlungen

Aus der bisherigen Untersuchung können die folgenden tabellarisch aufgeführten Handlungsempfehlungen abgeleitet werden.

Bereich Handlungsempfehlungen	**Handlungsempfehlungen**
IT-Strategie	Die IT-Strategien aller drei Gemeinden sollten mehr auf die Legislaturziele fokussieren und auch einen Beitrag dazu leisten.
Rolle der Informatik	Bei allen drei Gemeinden sollte überprüft werden, wohin sie sich strategisch entwickeln wollen. Ihre Rollen müssen danach entsprechend ausgerichtet und gelebt werden.

Angebot und Einkauf von IT-Dienstleistungen	Der IT der Gemeinde A ist zu empfehlen, den Einkauf von IT-Dienstleistungen zu standardisieren, damit sie Lieferanten und Consulting-Firmen unabhängiger wird. Auch bei Gemeinde B und C bestehen bezüglich IT-Dienstleistungen Verbesserungspotentiale, in dem diese beispielsweise quantitativ messbar gemacht und laufend optimiert werden.
Kundenbeziehungen, Zusammenarbeit mit den Fachabteilungen und Kommunikation	Bei allen drei IT-Abteilungen können die Beziehungen zu den Kunden optimiert werden.
Operative Gremien	Die IT der Gemeinde A sollte möglichst alle Direktionen bei Informatik-bezogenen Sitzungen einbeziehen. Der IT der Gemeinde B wird nahegelegt, auf der Durchführung des Informatikausschusses zu bestehen. Zudem sollte sie sich dafür einsetzen, dass IT-Themen auch in der Geschäftsleitung behandelt werden. Ideal wäre, wenn sich die Informatik zu einem Querschnittsbereich über alle Direktionen entwickeln könnte. Gemeinde C könnte seine IT-Gremien weiter ausbauen und den Austausch optimieren.

Abbildung 8: Handlungsempfehlungen aus der BIA-Maturitätsprüfung zuhanden der drei Gemeinden.

7.3 Ausblick

Das eher ernüchternde Ergebnis einer tiefen bis mittleren Reife des BIA bei den untersuchten Berner Gemeinden ist nicht weiter überraschend. Bereits im Vorfeld der Interviewphase hat sich gezeigt, dass die eine oder andere Gemeinde kein Abgleich zwischen Legislaturzielen und IT-Strategie vornimmt. Es stellt sich somit die Frage, warum BIA bei den Gemeinden (noch) nicht sehr verbreitet ist. Besteht zu wenig Druck dazu oder wird es einfach nicht benötigt? Sind IT-Strategien mit einem Bezug zu den Legislaturzielen nicht gewollt oder werden sie nicht gebraucht? Wird die Informatik von den Direktionen und dem Gemeinderat tatsächlich nur als Enabler betrachtet, dem nicht mehr Bedeutung beigemessen werden muss? Und wird dies in Zukunft auch so bleiben? Es könnte sein, dass die IT der Gemeinden kein kritischer Erfolgsfaktor ist oder die Gemeinden diesen noch nicht erkannt haben. Jedenfalls sind dies Fragen, die mit einer Nachfolgestudie zu untersuchen sind. Interessant wäre insbesondere, zusätzlich die Sicht von Direktionen, Gemeindepräsidenten und Ratsvertretern einzubringen. Diese Studie ging mehrheitlich von der Befragung der IT aus. Bei den befragten IT-Leitenden scheint das Themas BIA jedenfalls von Bedeutung zu sein. Es wäre sogar mehr Abgleich und Abstimmung zwischen der IT und den Abteilungen wünschenswert. Die Informatik wird definitiv an Wichtigkeit gewinnen. Gerade im Bereich von E-Government wird sich in den nächsten Jahren einiges ändern, wodurch viele Neuerungen (Netzwerk, Server, etc.) zu erwarten sind. Durch Gemeindefusionen werden Komplexität und Abhängigkeiten zwischen Geschäft und IT und innerhalb der IT zunehmen. Es wird vermehrt Fachleute brauchen. Aus NPM-Sicht muss auch die IT gegenüber den Bürgern unter wirtschaftlichen Gesichtspunkten eine gute Leistung erbringen. Alle diese Aspekte deuten darauf hin, dass auch das BIA künftig in Gemeinden immer wichtiger werden wird. Jedoch muss auch berücksichtigt werden, dass bei den Gemeinden teilweise andere Bedürfnisse

im Vordergrund stehen. Weiter spielen auch die vorhandenen Ressourcen (personelle, zeitliche und finanzielle Kapazität) eine wesentliche Rolle. In einem Fall gilt: Würde die Informatik beispielsweise proaktiv auf die Direktionen zugehen, würden Wünsche und Anforderungen geäußert, denen aber mit den vorhandenen Ressourcen nicht nachgekommen werden könnte. Es kann also gesagt werden, dass bei den Gemeinden die Informatik z.T. noch zu stark auf die Technologien fokussiert, anstatt sich für die Optimierung von Geschäftsprozessen und Unterstützung der Fachabteilungen einzusetzen, um so einen optimalen Strategiebeitrag und damit Wertschöpfung für die Gemeinde zu leisten. Aus diesen Gründen ist zurzeit offen, wie sich das Thema BIA bei öffentlichen Verwaltungen wie beispielsweise Gemeinden in Zukunft entwickeln wird. Jedoch ist zu hoffen, dass das Thema durch anstehende Neuerungen wie E-Government oder E-Voting mehr Bedeutung erhält und entsprechend angegangen wird. Auch aus Sicht von Wirtschaftlichkeit und Nachhaltigkeit sind strategische und operative Abgleiche zwischen Legislaturzielen und IT-Strategien, sofern diese dann auch existieren, sicherlich ertragreich.

Literaturverzeichnis

[AJP04] Avison, D.; Jones, J.; Powell, P.; Wilson, D.: Using and validating the strategic alignment model. In: Journal of Strategic Information Systems 13 (2004), S. 223–246.

[BEH10] Bashiri, I.; Engels, C.; Heinzelmann, M. (2010): Strategic Alignment. Berlin: Springer Verlag.

[BWF07] Beimborn, D.; Wagner, H.-T.; Franke, J.; Weitzel, T. (2007): The Influence of Alignment on the Post-Implementation Success of a Core Banking. Frankfurt: IEE Computer Science.

[Be06] Beveridge, C. (2006): Aligning IT with Business Strategy. Manchester: The National Computing Centre Ltd.

[BET13] Bigler, N.; Enkerli, S.; Topfel, M. (2013): Business IT Alignment bei Gemeinden, Abschlussbericht, Master Wirtschaftsinformatik, aF+E Fälle & Übungen, Berner Fachhochschule, Bern.

[CKA05] Campell, B.; Kay, R.; Avison, D.: Strategic alignment: a practitioner's perspective. In: Journal of Enterprise Information Management 18 (2005) 6, S. 653 – 664.

[CMR14] Christen, P.; Maag, S.; Regli, A. (2014): Qualitative Untersuchung des Business-IT-Alignments bei grossen Berner Gemeinden, aFuE-Arbeit Master Wirtschaftsinformatik, Berner Fachhochschule, Bern.

[CM11] CMMI Product Team (2011): CMMI® für Entwicklung, Version 1.3, SEI-sanctioned GERMAN translation of CMMI-DEV, V1.3, Prozessverbesserung für die Entwicklung besserer Produkte und Dienstleistungen.

[CMoJ] CMMI-Levels (ohne Jahr): Capability Maturity Model Integration (CMMI Levels). Abgerufen am 06.05.2014 von http://www.cmmilevels.com/.

[DG09] De Haes, S.; Grembergen, W.v.: An Exploratory Study into IT Governance Implementations and its Impact on Business/IT Alignment. In: Information Systems

Management 26 (2009) 2.

[GKB07] Gomber, P.; König, W.; Beimborn, D.; Franke, J.; Wagner, H.-T.; Weitzel, T. (2007): Kritische Erfolgsfaktoren in Finanzprozessen – IT-Management und Alignment im Kreditprozess der 1.000 größten Banken in Deutschland.

[HD12] Hammerström, I.; Dyer , M. (2012): Prozesse und Werkzeuge für ein gelebtes Business-IT-Alignment. Zürich: AWK GROUP.

[HV93] Henderson, J. C.; Venkatraman, N. (1993): Strategic Alignment: Leveraging information technology for transforming organizations. IBM Systems Journal, Vol. 32, No. 1, S. 472-484.

[LB99] Luftman, J.; Brier, T.: Achieving and sustaining Business-IT Alignment. In: California Management Review 42 (1999) 1/Fall, S. 109-122.

[LPB99] Luftman, J.N.; Papp, R.; Brier, T.: Enablers and Inhibitors of Business-IT Alignment. In: Communications of AIS 1 (1999) Article 11, S. 1-33.

[Lu00] Luftman, J. (2000). Assessing Business-IT Alignment Maturity. Communications of AIS, Volume 4, Article 14.

[Ni04] Nickels, D. W. (2004): IT-Business alignment: what we know that we still don't know. Memphis: University of Memphis.

[SC01] Sabherwal, R.; Chan, Y.E.: Alignment Between Business and IS Strategies: A Study of Prospectors, Analyzers, and Defenders. In: Information Systems Research 12 (2001) 1, S. 11-23.

[Vi11] Vimentis (2011): Vimentis. Auf: http://www.vimentis.ch/d/lexikon/498/Legislatur.html (Aufruf per 17.05.2014).

[WEB14] Walser, K.; Enkerli, S.; Bigler, N.; Topfel, M. (2014): Business-IT aligment in municipalities – The Swiss case. In: Proceedings of the 8th International Conference on Methodologies, Technologies and Tools Enabling e-Government (MeTTeG14), Udine, Italy – September 25th -26th, 2014, S. 53-66.

[WBS14] Weitzel, T.; Beimborn, D.; Schlosser, F. (2014): Universität Bamberg, Information Systems and Services, IT Business Alignment. Abgerufen am 30.05.2014 von http://www.uni-bamberg.de/isdl/leistungen/transfer/it-business-alignment/.

[Yi09] Yin, R. K. (2009): Case Study Research, Design and Methods (Fourth Edition). Inc. Thousand Oaks: Sage Publications.

8 Anhang

Link zum Dropbox-Dokument mit den ausführlichen Forschungsresultaten:

https://www.dropbox.com/s/9uxyprlv3pi65qb/Ausf_Forschungserg_BIA_Gemeind_CaseStudies%20-%20Kopie.pdf?dl=0

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Workshop „Business Process Management im Öffentlichen Sektor“ (BPMÖS)

Workshop „Business Process Management im Öffentlichen Sektor“ (BPMÖS)

Abstract zum Workshop
Der Workshop beschäftigt sich mit dem Thema Business Process Management (BPM) im Öffentlichen Sektor. Darunter werden die Branchen Öffentliche Verwaltung, Gesundheitswesen und Non-Profit-Unternehmen des Öffentlichen Sektors gezählt.
In Folge eines doppelblinden Begutachtungsverfahrens resultierten zwei ausgewählte Beiträge zum Gegenstand des Workshops. Zudem wurden zwei Vorträge eingeladen.
Die erste eingeladene Keynote beschäftigt sich mit dem Themenbereich Process Mining, das auch im Öffentlichen Sektor erfolgreich eingesetzt wird. Der Vortrag fokussiert insbesondere die Anwendungsmöglichkeiten und Positionierung von Process Mining im allgemeinen Themengebiet Geschäftsprozessmanagement.
Der zweite eingeladene Vortrag beschäftigt sich mit der Thematik des Prozessmanagements bei der Deutschen Bundesbahn. Dabei steht die Frage im Vordergrund, wie im Rahmen des Bundesnaturschutzgesetzes (BNatSchG) umfangreiche Dokumentations- und Reportingpflichten u.a. mittels Geschäftsprozessmanagement unterstützt wurden. Dazu musste ein flexibles und nachhaltiges IT-System geschaffen werden, welches die komplexen Prozesse rund um naturschutzspezifische Kompensationsmaßnahmen aktiv steuert und interne wie externe Anwender bei der umfangreichen Datenerhebung und Berichtserstellung aktiv unterstützt. Der Vortrag zeigt, wie mittels BPM und SCRUM (ein Vorgehensmodell) die agile Konzeption, Entwicklung und Einführung von eines Fach-Informationssystem-Naturschutz und Kompensation mit Kurznamen FINK verlaufen ist.
Der Beitrag von Holzmüller-Laue/Häger beschäftigt sich vor dem Hintergrund des E-Government-Projektes Föderales Informationsmanagement (FIM) des IT-Planungsrates mit der Frage der Evaluation einer Methodik zur standardisierten Abbildung von Verwaltungsprozessen im Bundesland Mecklenburg-Vorpommern. Dabei wurde der FIM-Ansatz in Richtung eines Kooperativen Informationsmanagements (KIM) weiter entwickelt. Der Beitrag beleuchtet dabei die auf verschiedenen Prozessoptimierungsprojekten basierende Evaluation der FIM/KIM-Standardisierungsvorhaben für den Baustein Prozesse innerhalb von FIM.
Der Beitrag von Enkerli/Walser thematisiert die Entwicklung eines Prozessreferenzmo-

dells für Kommunen sowohl auf Basis des ARIS-Modells als auch auf Grundlagen im Bereich Standardisierung des Geschäftsprozessmanagements ausgehend von eCH (Verein zur Entwicklung von E-Government-Standards in der Schweiz; www.ech.ch). Dabei resultiert ein Rahmenwerk zur standardisierten Entwicklung von Geschäftsprozessreferenzmodellen für Kommunen, das de facto weltweit (auch auf anderen föderalen Ebenen) eingesetzt werden kann, sofern gewisse Voraussetzungen im Sinne von Leistungskatalogen etc. vorhanden sind. Die Dimensionen des entsprechenden Referenzmodells umfassen neben dem Thema Prozesse auch ein internes Kontrollsystem (IKS), Informationsflüsse, RACI-Charts (Darstellungsform zur Analyse von Verantwortlichkeiten), Organigramme, Formulare, relevante Dokumente (Gesetze), betroffene IT-Systeme und mehr.
Vor dem Hintergrund u.a. der Tatsache, dass unlängst der Betrieb des ebenfalls vom Bundesministerium des Inneren geförderten E-Government-Projekts für eine Prozessaustauschplattform (Nationale Prozessbibliothek (NPB)) nicht weiter gefördert wurde und der Betrieb damit eingestellt werden musste, werden im weiteren Verlauf des Workshops zuhänden der GI-Fachgruppe Verwaltungsinformatik Empfehlungen entwickelt, wie das Geschäftsprozessmanagement in der öffentlichen Verwaltung mehr Relevanz erhält, was Förderungs- und Hinderungsfaktoren dafür sind sowie welche Maßnahmen zu ergreifen sind, um auch Automatisierungen von Geschäftsprozessen in Fachverfahren stärker zu fördern.

Organisationskomitee des Workshops

Prof. Dr. Konrad Walser
Professor für Wirtschaftsinformatik und E-Government

Berner Fachhochschule
E-Government-Institut
Brückenstrasse 73
CH-3005 Bern

E-Mail: konrad.walser@bfh.ch

Prof. Dr. André Göbel
Vertr. Lehrgebiet Verwaltungsmanagement und Wirtschaftsförderung

Hochschule Harz
FB Verwaltungswissenschaften
Am Domplatz 16
D-38820 Halberstadt

E-Mail: agoebel@hs-harz.de

Prof. Dr. Dagmar Lück-Schneider
Lehrgebiet Verwaltungsinformatik

Hochschule für Wirtschaft und Recht
Berlin (HWR Berlin)
Fachbereich Allgemeine Verwaltung
Alt-Friedrichsfelde 60
D-10315 Berlin

E-Mail: Dagmar.Lueck-Schneider@hwr-berlin.de

Programmkomitee des Workshops

- Dr. Norbert Ahrend – Nationale Prozessbibliothek/Humboldt Universität Berlin
- Prof. Dr. Thomas Allweyer – Hochschule Kaiserslautern
- Prof. Dr. Roland Blomer – UMIT, Tirol
- Prof. Dr. Michael Breidung – Eigenbetrieb IT der Stadt Dresden
- Prof. Dr. Andreas Breiter – ifib / Universität Bremen
- Julian Detzel – Capgemini, Berlin
- Georg Gelhausen – Bundesministerium des Inneren BMI, Bonn
- Prof. Dr. Knut Hinkelmann – Fachhochschule Nordwestschweiz, Olten
- Prof. Dr. Georg Rainer Hofmann – Hochschule Aschaffenburg
- Prof. Dr. Andreas Meier – DIUF, Universität Fribourg
- Prof. Dr. Jan Mendling – Wirtschaftsuniversität Wien
- Prof. Dr. Maria Wimmer – Universität Koblenz-Landau
- Dr. Michael Räckers – European Research Center for Information Systems ERCIS, Westfälische Wilhelms-Universität Münster
- Prof. Dr. Ulrich Reimer, Fachhochschule St. Gallen FHSG
- Prof. Dr. Tino Schuppan – IFG.CC, Potsdam
- Prof. Dr. Barbara Thönssen – Fachhochschule Nordwestschweiz FHNW, Olten
- Prof. Dr. Jan vom Brocke – Universität Liechtenstein
- Prof. Dr. Jörn von Lucke – Zeppelin-Universität, Friedrichshafen
- Prof. Dr. Mathias Weske, Hasso Plattner Institut an der Universität Potsdam
- Dr. Martin Wind, Abteilungsleiter, Senatorin für Soziales, Kinder, Jugend und Frauen, Stadt Bremen
- Dr. Petra Wolf – Fakultät für Informatik, Technische Universität München

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Evaluation einer Methodik zur standardisierten Abbildung von Verwaltungsprozessen

Silke Holzmüller-Laue[1] und Kati Häger[2]

Abstract: Prozessmanagement (GPM) stellt zunehmend in der öffentlichen Verwaltung ein Instrument von wachsender Bedeutung dar, um den aktuellen Herausforderungen gerecht zu werden. Das auf Bundesebene initiierte Projekt „Föderales Informationsmanagement (FIM)“ des IT-Planungsrates schlägt für Verwaltungsverfahren einen Standardisierungsrahmen der Informationsbausteine Leistungen, Formulare und Prozesse vor. Um die Durchsetzung des Prozessmanagement in der Verwaltung auf eine breite Basis zu stellen, wurden die FIM-Ansätze in Mecklenburg-Vorpommern (M-V) für die ausführende Ebene im Rahmen des „Kooperativen Informationsmanagement (KIM.MV)“ weiterentwickelt. Dieser Beitrag beschreibt die auf verschiedenen Prozessoptimierungsprojekten basierende Evaluation des KIM.MV-Standardisierungsvorhabens für den Baustein Prozesse.

Keywords: Evaluation, Föderales Informationsmanagement, Kooperatives Informationsmanagement, Prozesse, Prozessmanagement, öffentliche Verwaltung.

1 Einleitung

Die demografische Entwicklung, eine angespannte Haushaltslage, gestiegene Kundenerwartungen aber auch die organisatorisch-technische Entwicklung machen eine Überprüfung der Effizienz von Verwaltungsabläufen erforderlich. Prozessmanagement bietet dazu eine standardisierte Methode, die bereits seit Jahren in der Industrie und Privatwirtschaft etabliert ist und zunehmend auch in der öffentlichen Verwaltung an Bedeutung gewinnt. Ziel der Prozessorientierung ist es, durch Verbesserung der Prozessabläufe die Wirtschaftlichkeit des Verwaltungshandelns bei einer gleichbleibenden bzw. verbesserten Qualität der angebotenen Dienstleistungen zu erhöhen, um so z. B. die Zufriedenheit der Verwaltungskunden (Bürger, Unternehmen, Verwaltungsorganisationen) zu steigern und schneller auf sich ändernde Rahmenbedingungen reagieren zu können.

In einer prozessorientierten Verwaltungsorganisation stehen die Ergebnisse des Verwaltungshandelns und die damit verbundenen Abläufe, welche als Prozesse verstanden werden, im Vordergrund. Die Einführung von Prozessmanagement in der öffentlichen Verwaltung ist eine organisatorische Herausforderung und erfordert eine

[1] DVZ Datenverarbeitungszentrum Mecklenburg-Vorpommern GmbH, Organisations- und Prozessberatung, Lübecker Straße 283, 19059 Schwerin, s.holzmueller-laue@dvz-mv.de
[2] DVZ Datenverarbeitungszentrum Mecklenburg-Vorpommern GmbH, Organisations- und Prozessberatung, Lübecker Straße 283, 19059 Schwerin, k.haeger@dvz-mv.de

Neuausrichtung des Verwaltungshandelns. Sie vollzieht den Paradigmenwechsel von der herkömmlichen aufgaben- und funktionsorientierten Organisation zu einer prozessorientierten Organisation, die Leistungen bereichsübergreifend, auch über Behördengrenzen hinweg, erbringt. Mit dem veränderten Verständnis, Verwaltungsleistungen als Dienstleistungen für Bürger, Unternehmen oder andere Verwaltungseinheiten zu sehen, rückt der Prozess der organisationsübergreifenden Leistungserstellung in den Fokus (Ende-zu-Ende-Ansatz). Verwaltungskunden sind Prozessbeteiligte, die stärker in die Prozesse integriert werden müssen als bisher, um die gesetzten Ziele (E-Government-Gesetz [EGovG]) zu erreichen. Prozessmanagement bietet hierbei Methoden und Werkzeuge, um Prozesse zu erheben, zu dokumentieren, zu gestalten, zu optimieren, auszuführen, zu steuern und stetig zu verbessern. Trotz vielfältiger Bestrebungen und Initiativen wird Prozessmanagement meist nur in ausgewählten Bereichen der Verwaltungen eingesetzt und ist vor allem in Landesbehörden kaum etabliert [PWC14].

Das Projekt „Föderales Informationsmanagement (FIM)“ des IT-Planungsrates hat sich zum Ziel gesetzt, die für Verwaltungsverfahren relevanten Informationen bzgl. Leistungen, Prozesse und Formulare zwischen Bund, Ländern und Kommunen auf freiwilliger Basis zu harmonisieren [FIM13]. Das FIM-Standardisierungskonzept soll die Qualität der Informationen erhöhen sowie den redaktionellen Aufwand deutlich reduzieren. Die Etablierung einer gemeinsamen Infrastruktur auf fachlich-redaktioneller und organisatorischer Ebene stellt ein weiteres, wesentliches Projektziel dar. [ITPR14]

Unter Federführung des Ministeriums für Inneres und Sport M-V wurde der FIM-Standardisierungsansatz aufgegriffen und in Empfehlungen für eine Vorgehensweise bei der Einführung und Weiterentwicklung von Prozessmanagement in den Verwaltungen M-V integriert. In einem kooperativen Projekt ist die FIM-Methodik für die Vollzugsebene weiterentwickelt worden, um durch Standardisierung das Prozessmanagement in den öffentlichen Verwaltungen auf eine breite Basis zu stellen. [PzMMV]

Dieser Beitrag beschreibt die auf verschiedenen Prozessoptimierungsprojekten basierende Evaluation dieser Methodik zur standardisierten Abbildung von Verwaltungsprozessen aus M-V.

2 Prozessmanagement in der öffentlichen Verwaltung

Die von pwc und IMTP 2014 herausgegebene Studie [PWC14] analysiert den aktuellen Stand der öffentlichen Verwaltung zum Prozessmanagement. Bescheinigt wird der Verwaltung durch die Studie, die Bedeutung von Prozessmanagement erkannt zu haben (80,4% setzen Prozessmanagement ein), aber es bisher nur in punktuellen, kurzfristigen Maßnahmen umzusetzen. Dabei wird der aktuelle Stand der methodischen Umsetzung als sehr heterogen beschrieben, z. B. bzgl. der eingesetzten Notation oder der Dokumentation (nur 16,8% nutzen eine Bibliothek, ansonsten eher papierbasiert). Heterogenität verhindert Vergleichbarkeit, Nachnutzbarkeit sowie eine organisations- und ebenenübergreifende Harmonisierung.

Es sind Instrumente erforderlich, welche die existierenden Ergebnisse aus Optimierungsprojekten des operativen Prozessmanagements in einen ganzheitlichen strategischen Ansatz überführen, um langfristig und kontinuierlich den genannten Herausforderungen begegnen zu können. Maßnahmen, welche die Kooperation, den Wissensaustausch und die behördenübergreifende Harmonisierung unterstützen, sind besonders hilfreich. Zusätzlich wird in der Studie betont, dass eine Verankerung des Prozessmanagements im Behördenmanagement für eine erfolgreiche Etablierung unabdingbar ist.

2.1 Instrumente

Neben bewährten allgemeinen GPM-Instrumenten wurden andere auf die speziellen Anforderungen und Rahmenbedingungen der öffentlichen Verwaltung zugeschnitten.

1. Verwaltungsspezifische Prozessnotationen: Neben den allgemeinen Prozessnotationen wie z. B. Ereignisgesteuerte Prozessketten (EPK), Business Process Model and Notation (BPMN) oder Unified Modeling Language (UML) wurden für die öffentliche Verwaltung spezielle Notationen entwickelt, die besonders die fachliche Modellierung des dokumenten- und informationsbezogenen Verwaltungshandelns erleichtern sollen.

- PICTURE ist eine Methode zur Modellierung und Beschreibung von Verwaltungsabläufen, deren Kern 24 vordefinierte Prozessbausteine bilden [Becker07]. Das Bausteinkonzept zielt auf Standardisierung, wodurch insbesondere Vergleichbarkeit und Analysierbarkeit erreicht werden sollen. Der hohe Abstraktionsgrad einiger Bausteine sowie der Verzicht auf explizite Verzweigungen führt zu relativ groben Prozessbeschreibungen, die für eine technische Umsetzung nicht geeignet sind. Obwohl es sich um eine auf einen Anbieter beschränkte proprietäre Notation ohne offenes Austauschformat handelt, ist diese Methode auf Grund ihrer Einfachheit und leichten Verständlichkeit gerade im kommunalen Bereich gut verbreitet. [NFIM13]

- Der Kommunale **Fa**ch**mo**dellierungs**s**tandard FaMoS wurde speziell für die Modellierung kommunaler Prozesse entwickelt. Dabei wird ein eingeschränktes Symbolset des Notationsstandards BPMN um zusätzliche an die Zielgruppe angepasste Symbole erweitert. Die große Anzahl von Symbolen führt zu sehr detaillierten Modellen, die zeitaufwendig zu erstellen sind [HoLa10]. FaMoS ist auf die Darstellung der organisatorischen Abläufe ohne Berücksichtigung von Kommunikationswegen beschränkt [NFIM13]. Obwohl FaMoS nur von wenigen Tools unterstützt wird (kein offenes Austauschformat), wird auch diese Notation im kommunalen Bereich relativ oft eingesetzt.

2. Austauschplattformen für die öffentliche Verwaltung

   Prozessbibliotheken für die öffentliche Verwaltung sind ein wichtiges Instrument für organisationsübergreifende Kollaboration und stoßen vor allem bei kommunalen

Verwaltungen auf Interesse. Sie werden primär zum Erfahrungsaustausch genutzt, eröffnen aber auch Nachnutzungseffekte für Prozessmodelle. Zu nennen sind hier vor allem das Prozessregister der Kommunale Gemeinschaftsstelle für Verwaltungsmanagement (KGSt) oder das des Freistaates Sachsen.

In der öffentlichen Verwaltung der Schweiz setzt man auf eine Standardisierung der GPM-Vorgehensweise, um die elektronische Zusammenarbeit („Vernetzte Verwaltung“) zwischen Behörden und von Behörden mit Privaten, Unternehmen und Organisationen zu erleichtern (eCH-Standard). Es werden Rahmenkonzepte, Standards und Hilfsmittel zum GPM bereitgestellt. Das eCH-Rahmenkonzept [eCH-0138] definiert die einheitliche Beschreibung und strukturierte Dokumentation von Aufgaben, Leistungen, Prozessen und Zugangsstrukturen der öffentlichen Verwaltung der Schweiz. [AhWa14]

2.2 Einführung Prozessmanagement in der öffentlichen Verwaltung von M-V

In M-V wurden in den letzten Jahren besondere Anstrengungen unternommen, um Prozessmanagement in der öffentlichen Verwaltung zu etablieren und damit einen Grundstein für ein effizientes und bürgernahes Verwaltungshandeln zu bilden [DPMV].

Das Büro kooperatives E-Government (BKE) stellte 2012 ein Methodenhandbuch (inkl. der Themen Prozesslandkarte, Rollenkonzept, Prozesslebenszyklus, Vorgehensmodell Prozessoptimierung) für die Einführung von Prozessmanagement in der öffentlichen Verwaltung in M-V zur Verfügung. In einem Folgeprojekt wurde die Praxistauglichkeit der empfohlenen Vorgehensweise von verschiedenen Behörden validiert, indem Prozessoptimierungsprojekte kooperativ durchgeführt wurden [WIWI14]. Schulungen und Coaching unterstützten die Behörden zudem zur eigenständigen Fortführung von Prozessmanagement über das Projekt hinaus. Ausgehend von einem zunächst offenen Ansatz ohne Vorgabe von Modellierungssprache oder -werkzeug wurde im Ergebnis BPMN 2.0 als gemeinsame Sprache für Prozesse festgelegt, um die Zusammenarbeit zu erleichtern und ein arbeitsteiliges Erstellen der Prozessmodelle zu ermöglichen. Des Weiteren wurde der Wunsch nach einer gemeinsamen Prozessmanagement-Software geäußert, so dass daraufhin das Folgeprojekt zum Aufbau einer gemeinsamen Prozessmanagement-Infrastruktur initiiert wurde. Den Verwaltungen in M-V steht damit ab 2016 ein weiteres wichtiges Instrument zur Verfügung, um Prozessmanagement nachhaltig und effizient realisieren zu können.

3 Föderales Informationsmanagement (FIM)

Das Projekt FIM (nachfolgend „FIM“ genannt) betrachtet die drei Informationsbausteine Leistungen, Formulare (Daten) und Prozesse sowie ihre Interaktionen, um Synergien bei der Ausführung von Verwaltungsleistungen auf allen föderalen Ebenen zu erschließen. FIM selbst steuert oder koordiniert nicht einzelne E-Government-Vorhaben, sondern gewährleistet Transparenz, Zusammenarbeit und Transaktion zwischen diesen.

Inhaltlich beschäftigt sich FIM mit der standardisierten Abbildung von bundesgesetzlichen Regelungen, den sogenannten Stamminformationen. Diese Grundidee basiert auf der Tatsache, dass 80% der Leistungen der öffentlichen Verwaltung auf Bundesgesetzgebung beruhen, und sich somit ein erheblicher Effizienzgewinn erzielen lässt, wenn die Informationen Top-down bereitgestellt werden.

Der **Baustein Formulare** fokussiert den standardisierten Austausch von formular- und dokumentenbasierten Informationen. Formulare ermöglichen der Verwaltung strukturiert und standardisiert Informationen zu erfassen (Antragsformulare) sowie bereitzustellen (Bescheide). FIM unterscheidet zwischen den Input- und Output-Dokumenten, die einen verwaltungsinternen oder verwaltungsexternen Ursprung haben. Steckbriefe zu verwaltungsinternen Formularen können durch Stammformulare (Zusammenstellung von Formularfeldern und Regeln) detailliert beschrieben werden. Ein wesentlicher Erfolgsfaktor und Treiber bei der Etablierung des bundesweiten Informationsmanagements ist auch die Verknüpfung der Formulardaten mit den Prozessinformationen.

Der **Baustein Leistungen** greift für die Definition und Beschreibung von Leistungen der öffentlichen Verwaltung auf die Ergebnisse des E-Government-Vorhabens LeiKa zurück, in dem der bereits etablierte Leistungskatalog (LeiKa) entwickelt wurde. Durch die einheitliche Zuordnung von verbindlichen LeiKa-Schlüsseln werden Leistungen im gesamten föderalen Verwaltungsgebiet vergleichbar. Die Leistungsbeschreibungen haben eine für alle föderalen Ebenen festgelegte fachliche Struktur und einheitliche Beschreibungsfelder, so dass sie auch inhaltlich gleich behandelt werden können.

Ein Prozess beschreibt die Aktivitäten, die zur Erbringung einer Leistung notwendig sind. Die eindeutige Zuordnung eines LeiKa-Schlüssel zum Prozess gibt den Ordnungsrahmen vor. Unterschiedlichen Informationsbedarfen Rechnung tragend, sieht FIM im **Baustein Prozesse** verschiedene Detaillierungsstufen (DS) zur Abbildung von Verwaltungsprozessen vor. Ausgehend von einem Prozesssteckbrief (DS1 - einfache Beschreibung eines Prozesses in strukturierter Textform) können Prozessinformationen eher abstrakt als Stammprozess (DS2) aber auch sehr feingranular auf der Ebene des Vollzugs (DS3) dargestellt werden. Der Stammprozess bildet mittels acht normierter, typisierter Verwaltungsaktivitäten, den sogenannten Referenzaktivitätengruppen (RAG), die allgemeingültigen Zusammenhänge basierend auf gültigen Rechtsnormen ab (Details in [FIM13]). Ein Stammprozess kann durch beliebig viele Lokalprozesse (DS3) detailliert werden, die aber über das Bundesprojekt nicht reglementiert werden. Der aus dem Gesetz abgeleitete einheitliche Prozesszuschnitt soll eine sinnvolle und effektive Nachnutzung sichern.

4 Kooperatives Informationsmanagement (KIM.MV)

Bisher waren kommunale Verwaltungstätigkeiten in den rund 11.000 Kommunen Deutschlands auf die lokale Aufgabenerfüllung ausgerichtet. Die immer leistungsfähigere Informations- und Kommunikationstechnologie ermöglicht, diese Verwaltungs-

tätigkeiten ebenenübergreifend zu standardisieren und miteinander zu vernetzen. Daraus resultierende Nachnutzungseffekte vermeiden Doppelarbeit und unterstützen einen Vergleich und Erfahrungsaustausch. Deshalb wird für die Verwaltungen in M-V der FIM-Gedanke im Kontext „Kooperatives Informationsmanagement (KIM.MV)“ auf der DS3 weiterentwickelt. Im Wesentlichen verfolgt der **KIM.MV-Baustein Prozesse** folgende Ziele:

1. Förderung der verwaltungsübergreifenden Kooperation durch einheitliches Prozessverständnis
2. Verbesserung der Nachnutzungsmöglichkeit von Prozessmodellen
3. Verbesserung der Vergleichbarkeit von Prozessen mit Auswertungen von Kennzahlen, Prozessreporting
4. Verbesserung der Auswertbarkeit bzgl. Schnittstellen, Medienbrüchen, verwendeten Prozessaktivitäten, ähnlichen Prozesssequenzen
5. Minimaler Aufwand bei der Erstellung und Pflege
6. Überprüfung der Rechtskonformität von Verwaltungshandeln
7. Entwicklung von Referenzprozessen zur sukzessiven Durchsetzung von (IT-) Standards (über Fachlichkeiten hinweg)
8. Standardisierung von Schnittstellen zwischen Partnerprozessen

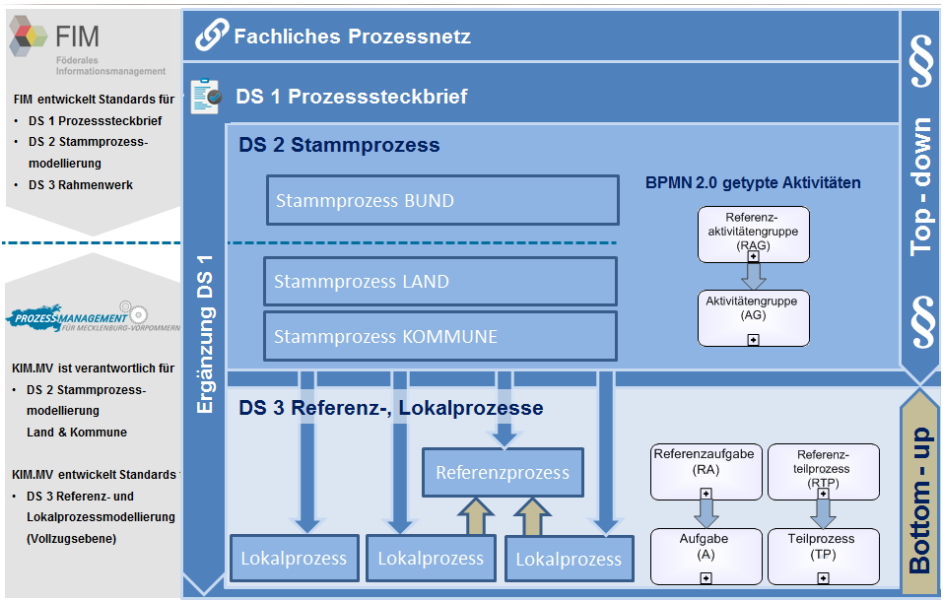


Abbildung 1 Erweiterung der FIM Methodik in die Vollzugsebene (KIM.MV)

Wie aus Abbildung 1 ersichtlich wird eine übergeordnete Strukturierungsebene „Fachliches Prozessnetz“ eingeführt. Das Prozessnetz enthält alle zusammenhängenden Prozesse des relevanten Fachthemas. Dabei wird zwischen den durch das Fachgesetz definierten Prozessen (Fachprozesse) und weiteren auf die Leistung einflussnehmenden Prozessen (Satellitenprozesse) unterschieden.

Ein Lokalprozess in der DS3 (Vollzugsebene) adaptiert den im Stammprozess beschriebenen Ablauf auf die Lokalspezifika einer ausführenden Organisation. Die Lokalprozesse beinhalten demnach neben den lokal eingesetzten IT-Fachverfahren bspw. auch Informationen zu den jeweils ausführenden Stellen. Durch die feingranulare Ausmodellierung der im Stammprozess beschriebenen Aktivitäten wird die Rechtskonformität auf der DS3 sichergestellt. Allerdings können sich in der Praxis Empfehlungen für Abläufe herauskristallisieren, die von der Rechtssetzung zwar abweichen, aber durchaus zu wertvollen Effizienzsteigerungen führen können. Diese Abweichungen werden im Lokalprozess farblich gekennzeichnet.

Ein Stammprozess kann in der Vollzugsebene unterschiedlich umgesetzt werden, so dass sich entlang der Verwaltungskaskade unterschiedliche Lokalprozesse ergeben können. Durch den Bezug zu den Rechtsnormen nimmt der Stammprozess stets eine stabile Ausgangsbasis für die Top-down Betrachtung ein. Aus Bottom-Up Sicht können Behörden aber auch mehrere Lokalprozesse zu einem gemeinsamen, übergreifenden Referenzprozess konsolidieren. Dieser ist methodisch der DS3 zuzuordnen.

Die DS3-Prozesse werden über 24 verwaltungsspezifische Typen von Aufgaben, den sogenannten Referenzaufgaben (RA), einheitlich strukturiert. Die Referenzaufgaben stellen Attribute bereit, um detailliertere Informationen zu dokumentieren und auszuwerten. Diese Attribute unterteilen sich in Kernattribute, die allen Aufgabentypen zugeordnet sind, und aufgabentypspezifische Attribute. Zur Bündelung von Aufgaben sind wiederum Teilprozesse erlaubt. Um jedoch eine einheitliche und somit vergleichbare Strukturierung der Prozesse auch auf der DS3 zu erreichen, dürfen die Teilprozesse nur den Typ einer RAG (DS2) einnehmen.

Auf der DS3 wird ein im Vergleich zur DS2-Modellierung erweitertes Set an BPMN 2.0 Elementen (z. B. weitere Typen von Verzweigungen, Zwischenereignisse) empfohlen, dessen Umfang mit den detaillierteren Informationsbedarfen der DS3 korreliert.

5 Evaluation

Das FIM-Konzept [FIM13] liegt derzeit noch nicht finalisiert vor, so dass eine Evaluation der KIM.MV-Erweiterung noch nicht auf einer abschließenden Beurteilung der praktischen Anwendbarkeit von FIM aufbauen kann. Deshalb wurde die im folgenden beschriebene Untersuchung auf den FIM-Baustein Prozesse ausgeweitet, auch wenn

den Autorinnen bewusst ist, dass sich wesentliche Vorteile und Potenziale der FIM-Methodik gerade aus dem Zusammenspiel der drei Bausteine ergeben. Die im Folgenden dargelegten Erkenntnisse entstanden in verschiedenen Prozessoptimierungsprojekten, welche die Autorinnen in der Landesverwaltung sowie im kommunalen Umfeld durchgeführt haben.

5.1 Vorgehensweise

Aus den jeweils relevanten Handlungsgrundlagen (gesetzliche Grundlagen, Verwaltungsvorschriften) wurde ein fachliches Prozessnetz abgeleitet. Die Prozesse wurden durch den Prozesssteckbrief beschrieben. Da derzeit kaum Bundes-Stammprozesse existieren, wurden entsprechende Landes-Stammprozesse basierend auf gültigen Rechtsnormen erstellt. Dazu wurden die durch die FIM-Methodik vorgeschlagenen standardisierten RAGs verwendet. Diese Stammprozesse wurden aus Bottom-Up-Sicht als Lokalprozesse ausdetailliert, welche die lokalen Spezifika und Rahmenbedingungen präzise abbilden. Grundlage hierfür waren lokale Handlungsgrundlagen, Formulare sowie die in Interviews erhobenen und anschließend optimierten Verwaltungsabläufe. Zur Modellierung der DS3-Prozesse wurden die in KIM.MV vorgeschlagenen 24 Referenzaufgaben verwendet.

5.2 Evaluation FIM - Prozesse

Die Evaluationskriterien wurden ausgehend von den drei primären FIM-Zielen definiert:

1. die verwaltungsebenen-, zugangskanal- und fachbereichsübergreifende Unterstützung der fachlichen, redaktionellen und technischen Interoperabilität von Leistungs-, Formular- und Prozessinformationen der öffentlichen Verwaltung,
2. die standardisierte Übersetzung von Rechtssprache in eine Vollzugssprache,
3. die Verringerung des redaktionellen Aufwands bei der Erstellung und Pflege dieser Informationen durch Umsetzung eines ebenenübergreifenden Redaktionskonzepts.

Ziel 1. Interoperabilität

1.1. Führt die FIM-Methodik zu einer homogenen Darstellung gleicher Sachverhalte?

Die acht vorgegebenen RAGs (Bausteinkonzept) tragen zu einer einheitlichen Strukturierung bei. Auch die Reduzierung des BPMN-Elementesets unterstützt das Ziel, im Ergebnis gleiche Modelle trotz Erstellung durch unterschiedliche Personen zu erhalten. Jedoch lassen Gesetzestexte mitunter Interpretationsspielraum zu, sind unterschiedlich abstrakt beschrieben, und geben die Reihenfolge der Aktivitäten nicht immer eindeutig vor, so dass unterschiedliche Darstellungen entstehen können. Für eine einheitliche Darstellung sind daher weitere Konventionen erforderlich, wie z. B. dass unabhängige Aktivitäten stets parallel zu modellieren sind.

1.2. Sind alle notwendigen Informationsbedarfe bedienbar?

Die gesetzlichen Vorgaben von z. B. IT-Systemen können durch die reduzierte FIM-BPMN nicht visualisiert, sondern nur über Attribute erfasst werden. Zukünftige Validierungen müssen zeigen, ob eine Erweiterung der FIM-BPMN ggf. notwendig ist.

Ziel 2. Übersetzung der Rechtssprache in eine Vollzugssprache

2.1. Ist die Ableitung des Prozesszuschnitts aus dem Gesetz praktikabel und eindeutig? Sind Rechtsnormen als Basis für eine Modellierung des Prozessablaufs geeignet?

Die Vorgabe des Prozesszuschnitts aus der DS1/DS2 ermöglicht einen interorganisationalen Vergleich auf Ebene der Lokalprozesse. Die Ableitung aus dem Gesetz ist prinzipiell praktikabel, hängt aber sehr von der Detaillierung der Rechtsnorm ab. Der LeiKa kann zur Beantwortung der Frage, wie viele Prozesse sich aus einem Gesetz ergeben, hinzugezogen werden. Leider hat sich dies in der Evaluation nur teilweise bewährt. Der LeiKa wurde in der Vergangenheit unter einer anderen Prämisse erstellt, so dass er aus FIM-Sicht zum heutigen Stand unvollständig ist. Eine kontinuierliche Weiterführung des LeiKa wird deshalb im Zusammenhang mit der FIM-Methodik empfohlen.

2.2. Definiert FIM Übersetzungsregeln von der Rechts- in die Vollzugssprache?

Im Feinkonzept [FIM13] finden sich darauf keine Antworten. Übersetzungsregeln werden benötigt, da sonst Experten- und Erfahrungswissen im jeweiligen Fachgebiet erforderlich ist. Hier müsste die Methodik geschärft werden. In der KIM.MV-Methodik wird dazu die Strukturierungsebene Prozessnetz mit Fach- und Satellitenprozessen eingeführt.

2.3. Sind die definierten RAGs für die Prozesse der öffentlichen Verwaltung geeignet?

Für das dokumenten- und informationsbezogene Verwaltungshandeln sind die sieben definierten RAG in den betrachteten Evaluierungsprozessen ausreichend, allerdings ist z. B. für alle Tathandlungen die Joker-RAG verwendet worden. Es wird deshalb vorgeschlagen, für praktische Tätigkeiten eine weitere RAG einzuführen.

2.4. Ist die FIM-BPMN für die Abbildung der Rechtsnormen geeignet?

Die Einschränkung des Elementeumfangs führt gerade in der Einführungsphase des Prozessmanagements zu besserer Verständlichkeit der Modelle. Interpretationsfehler werden vermieden und Verwaltungsmitarbeiter fühlen sich nicht überfordert. In den Evaluationsprozessen allerdings fehlten Zwischenereignisse (z. B. für gesetzlich definierten Fristen) sowie Datenspeicher (z. B. für IT-Systeme).

Ziel 3. Verringerung des redaktionellen Aufwands

3.1. Stabilität der Informationen

Die Prozesse bleiben über einen längeren Zeitraum valide, da durch die zugrunde

liegenden Rechtsnormen ein stabiles Bezugssystem gegeben ist. Die Prozesse ändern sich nur, wenn das Gesetz sich ändert. Das Attribut „gesetzliche Grundlagen“ ermöglicht eine Identifikation der von einer Gesetzesänderung betroffenen Prozesse bis in die Vollzugsebene. Dadurch wird die Gesetzesfolgenabschätzung erleichtert.

3.2. Allgemeingültigkeit und Übertragbarkeit

Die in den Stammprozesse enthaltenen gesetzlichen Grundlagen werden auf die Lokalprozesse übertragen und gelten für alle betroffenen Organisationen. Auch wenn in der Praxis aufgrund der fehlenden Bundesredaktion dieser Punkt noch nicht wie angedacht umgesetzt ist, konnte der Ansatz bestätigt werden, indem ein Stammprozess aus dem Bundesgesetz abgeleitet und in zwei Kommunen die Lokalprozesse erstellt wurden.

3.3. Effektivität durch Arbeitsteilung

Der gesetzesbasierte Prozesszuschnitt ermöglicht die arbeitsteilige Erstellung der Modelle bzw. eine Nachnutzung existierender Modelle. Eine strukturierte Ablage in einem recherchierbaren Prozessregister ist Voraussetzung für jegliche Art der Nachnutzung, aber derzeit noch nicht global verfügbar.

5.3 Evaluation KIM.MV-Baustein Prozesse

Auch für die Erweiterung der Methodik für die DS3 wurden die Prüfkriterien aus den Zielen (Abschnitt 4) abgeleitet.

Ziel 1. Förderung der verwaltungsübergreifenden Kooperation

1.1. Einheitliche Modellierungssprache

In der Validierung des Methodenhandbuchs „Prozessmanagement in M-V“ ergab sich als ein wichtiges Projektergebnis die Empfehlung für eine einheitliche Prozesssprache (BPMN 2.0), um das gegenseitige Verständnis zu verbessern und die „Übersetzungsaufwände“ zu minimieren.

1.2. Einführung von gemeinsamen Konventionen

Die KIM.MV-Methodik führt den Ansatz der Standardisierung durch typisierte Bausteine fort, in dem elementare Verwaltungsaufgaben in der Vollzugsebene durch 24 Typen klassifiziert werden. Die definierten Aufgabentypen haben denselben Detaillierungsgrad, so dass die Betrachtung der Aufgaben verwaltungsübergreifend auf konstanter und vergleichbarer „Flughöhe“ erfolgt. Damit können beliebige verwaltungsübergreifende Prozesse kooperativ durch das Zusammensetzen von standardisierten Prozess-Puzzleteilen entstehen, was die Effizienz von Kooperationen im Prozessmanagement entscheidend vorantreiben kann.

Ziel 2. Verbesserung der Nachnutzungsmöglichkeit von Prozessmodellen

2.1. Standardisierte Prozessbeschreibung

Durch die Festlegung auf eine gemeinsame Modellierungssprache mit einheitlichen Konventionen wird der Aufwand von Anpassungen bei nachgenutzten Prozessmodellen deutlich reduziert, da jeglicher sprachbedingter Transformationsaufwand entfällt.

2.2. Publizität und Prozessmanagement-Software

Eine Nachnutzung von Prozessmodellen durch andere Verwaltungen setzt neben einer Standardisierung vor allem voraus, dass die Prozessmodelle für alle Interessierten zugänglich sind. Für KIM.MV wird derzeit eine zentrale Prozessmanagement-Software mit einem gemeinsamen Repository als Austauschplattform beschafft. Die dadurch gewonnene Transparenz sowie die leichte Übernahme der Prozessmodelle für lokale Anpassungen wird dadurch enorm erleichtert.

Ziel 3. Verbesserung der Vergleichbarkeit von Prozessen und Auswertungen

3.1. Gleicher Prozesszuschnitt

Erst wenn Prozessgrenzen (Anfang und Ende) klar und einheitlich bestimmt werden, lassen sich Kennzahlen und Auswertungen von Prozessen vergleichen. Der KIM.MV-Baustein Prozesse konkretisiert FIM durch Festlegungen, wie sich ein Prozessnetz von fachlich zusammengehörenden Prozessen zusammensetzt.

3.2. Gleiche Strukturierung

Die im Stammprozess normierten Aktivitätengruppen liefern die Struktur für die instanziierten DS3-Vollzugsprozesse. Die Kennzahlen der DS3-Prozesse können zu den korrespondierenden Aktivitätengruppen des Stammprozesses aggregiert werden. Bundesweit lassen sich so die Kennzahlen der DS3-Vollzugsprozesse aktivitäten-gruppenscharf auf der DS2 vergleichen. Der KIM.MV-Baustein Prozesse stellt zudem weitere gleich-granulare Aufgabentypen zu einer einheitlichen vertikalen und horizontalen Strukturierung der DS3-Prozesse zur Verfügung. Teilprozesse sind erlaubt, sofern sie einem Typ einer RAG der DS2 entsprechen. Auf diese Weise ist auf der DS3 die Vergleichbarkeit von Verwaltungsaufgaben im Vollzug sichergestellt.

3.3. Gleiche Attribute

Die normierten Aktivitäten werden durch vordefinierte Attribute (z. B. Kennzahlen, Softwaresystem) näher beschrieben. Die Attribute sind automatisiert auswertbar und durch den gleichen Prozesszuschnitt und die gleiche Strukturierung der Prozesse vergleichbar.

Ziel 4. Verbesserung der Auswertbarkeit bzgl. Schnittstellen usw.

4.1. Normierte Typisierung von Aktivitäten

Durch die Visualisierung der Verwaltungsabläufe lassen sich im Prozessmodell sehr schnell Schnittstellen erkennen. Die Prozessanalyse kann jedoch zusätzlich durch automatisierte Prozessauswertungen und -reports unterstützt werden. Dies ist jedoch insbesondere von der Qualität der Attributdaten abhängig. Der KIM.MV-Baustein Prozesse unterstützt dies insbesondere durch die 24 Referenzaufgaben. Pflichtattribute pro Aufgabentyp garantieren die Auswertbarkeit. Bestimmte Verwaltungsaufgabentypen deuten auf Medienbrüche oder Schnittstellen hin. Durch die Typisierung von Aufgaben lassen sich zudem Wiederholungen von aufeinanderfolgenden Tätigkeiten leicht identifizieren und als typische Prozesssequenzen in anderen Prozessen wiederverwenden. Die Standardisierung von Prozesssequenzen ist damit auch eine Ausgangsbasis für die Durchführung als Shared Services.

Ziel 5. Minimaler Aufwand bei der Erstellung und Pflege

5.1. Effektivität bei Prozesserstellung

Durch die Übernahme der Stammprozesse aus der redaktionellen Kaskade wird der Aufwand zur Erstellung der DS3-Prozesse erheblich reduziert. Die Verwendung derselben Modellierungssprache (BPMN 2.0) durch FIM und KIM.MV vermeidet Übersetzungsaufwände. Mit BPMN 2.0 als DS3-Standard und einer gemeinsamen Prozessmanagement-Software unterstützt der KIM.MV-Baustein Prozesse den Austausch, die arbeitsteilige Erstellung sowie die Nachnutzung von Prozessen zwischen den Behörden.

5.2. Kooperation

Für eine flächendeckende Etablierung von Prozessmanagement ist die effiziente Kooperation zwischen Behörden unabdinglich. Im KIM.MV-Baustein Prozesse wird derzeit ein landesweites Organisationsmodell entwickelt, dass u.a. die Steuerung der Referenzprozesserstellung als zentrale Aufgabe verortet.

Ziel 6. Überprüfung der Rechtskonformität von Verwaltungshandeln

6.1. Praktikabilität

Durch die Instanziierung der DS3-Prozesse aus dem DS2-Stammprozess werden die zugrunde liegenden Rechtsnormen übernommen. Ein direkter Abgleich zwischen den in der Aktivitätengruppe hinterlegten Rechtsnormen mit dem im DS3-Prozess detailliert beschriebenen Ablauf auf der Vollzugsebene ermöglicht eine effiziente Überprüfung der Rechtskonformität des Verwaltungshandeln ohne aufwendiges Recherchieren. Des Weiteren kann beim Hinzuziehen des Stammprozesses bei der DS3-Prozesserstellung verhindert werden, dass Sachverhalte, die verbindlich in gesetzlichen Regelungen verankert sind, bei der Modellierung der DS3-Prozesse vergessen oder weggelassen werden.

Ziel 7. Entwicklung von Referenzprozessen

7.1. Vergleichbare Lokalprozesse

Der KIM.MV-Baustein Prozesse unterscheidet in der DS3 die Lokal- und die Referenzprozesse. Sofern ein Prozess für mehr als zwei Organisationen gilt, handelt es sich um einen Referenzprozess. Dieser kann Bottom-Up auf Grundlage fachlich vergleichbarer Lokalprozesse erstellt werden. Durch die im KIM.MV-Baustein Prozesse festgelegten Konventionen werden Lokalprozesse einheitlich beschrieben, so dass eine Abstraktion der verwaltungsübergreifend gleichen Aufgaben in Form eines Referenzprozesses effektiv vorgenommen werden kann.

7.2. Zentrale Steuerung und Kooperation

Die Priorisierung von zu erstellenden Referenzprozessen ist im Organisationsmodell verortet. Derzeit werden Vorschläge für Referenzprozesse zentral ausgewertet, deren Erstellung kooperativ und verwaltungsübergreifend erfolgen soll.

Ziel 8. Standardisierung von Schnittstellen zwischen Partnerprozessen

8.1. Normierung von Aktivitätentypen und Attributen, sowie Daten

Manche Aufgabentypen implizieren Schnittstellen zu anderen Prozessen, die durch Attribute näher beschrieben werden. Solche Schnittstellen können automatisch identifiziert werden. In Anlehnung an FIM werden auch in KIM.MV die Prozesse und Formulare enger verzahnt, so dass der Datenaustausch zwischen lokalen Partnerprozessen schneller identifiziert, besser analysiert und übergreifend standardisiert werden kann.

5.4 Ergebnisse

BPMN ist für die Modellierung von Verwaltungsprozessen gut geeignet. Sie ist übersichtlich, lesbar und erfüllt alle wesentlichen Informationsbedarfe der Verwaltung. Die Einschränkung der FIM-BPMN muss weiter überprüft werden, ggf. werden weitere Elemente zur Abbildung der Rechtsnormen benötigt. Durch das hohe Abstraktionsniveau sind FIM-Stammprozesse aber wiederum managementtauglich.

Der Top-Down-Ansatz gewährleistet einen gleichen Prozesszuschnitt und ermöglicht somit eine vielfältige Wiederverwendbarkeit der Prozessmodelle. Die einheitliche Prozessvisualisierung der DS2 erlaubt eine eindeutige Identifikation von Unterschieden in den Vollzugsprozessen der DS3. Die standardisierte Modellierung über die redaktionelle Kaskade reduziert den Aufwand für das Erstellen und die Pflege der Prozessmodelle erheblich. Kommunale Verwaltungen erwarten eine Aufwandsreduktion auf 40-50% durch die Nutzung der vorhandenen Stammprozesse (Schätzung).

In allen untersuchten Beispielprozessen zeigte sich, dass die Standardisierung zu einem homogenen Detailgrad und einer einheitlichen Strukturierung der Prozesse führt, was

Vergleichbarkeit und Nachnutzbarkeit ermöglicht. Die RAGs sind nur für dokumenten- und informationsbezogene Verwaltungstätigkeiten geeignet. Tathandlungen müssen durch die „Joker"-RAG abgebildet werden. Dies ist nicht ausreichend. Weiterhin sollte geprüft werden, ob die dokumentenbezogene Sichtweise auf das Verwaltungshandeln im Hinblick auf elektronischen Rechtsverkehr und E-Akte noch zeitgemäß ist oder einige RAGs einer Anpassung in Bezeichnung und Bedeutung erfordern.

Neben den Ergebnissen zu den genannten Prüfkriterien haben sich während der durchgeführten Untersuchungen weitere Erkenntnisse ergeben:

- Ein Soll-Prozess kann auch vom Stammprozess abweichen, wenn z. B. langfristig die gesetzlichen Grundlagen angepasst werden müssen.

- Es kann als Vorgehensweise empfohlen werden, die Ist-Prozesse über die RAGs der DS2 einheitlich zu strukturieren. Dies erleichtert anschließend den IST-SOLL-Vergleich sowie die Prüfung auf Gesetzeskonformität des Ist-Prozesses.

- Die Anwendung der FIM-Methodik eröffnet Wege zu weiteren Effizienzsteigerungen, z. B. wird dadurch eine effiziente Vollzugsaufwandsabschätzung zur quantitativen Bestimmung der Folgen von Rechtsvorschriften und deren vergleichender Bewertung ermöglicht. Eine langfristig anzustrebende Integration der Prozessmodellierung in das Gesetzgebungsverfahren schafft die bestmöglichen Voraussetzungen dazu. Wird direkt bei der Erstellung der Gesetzestexte gleichzeitig schon ein standardisiertes Prozessmodell erstellt, lässt sich die erforderliche Vollzugsaufwandsermittlung prozessbasiert und damit wesentlich effizienter durchführen. Neue Vorgehensweisen wie die arbeitsteilige Analyse durch mehrere ausführende Verwaltungen werden ermöglicht.

6 Zusammenfassung

Das Projekt FIM hat sich zum Ziel gesetzt, die redaktionellen und technischen Voraussetzungen für die koordinierte Erstellung und Pflege der Stamminformationen (einheitliche Beschreibung von Leistungen, Prozessen und Formulare) zu schaffen und bietet dafür einen gemeinsamen Standardisierungsrahmen. Allerdings stellen die unterschiedlichen Entwicklungsstände und –geschwindigkeiten in den FIM-Bausteinen derzeit noch eine wesentliche Herausforderung für deren Vernetzung dar.

Der Erfolgsfaktor „Kooperation" liegt weiterhin im Ermessen der Verwaltung, jedoch unterstützt die FIM/KIM.MV-Methodik durch die Standardisierung und durch gültige Rechtsnormen als gemeinsamen Ausgangspunkt effektiv die organisationsübergreifende Zusammenarbeit. Die FIM/KIM.MV-Methodik eröffnet die Möglichkeit, die von einer Gesetzesänderung betroffenen Prozesse einfach und schnell zu identifizieren. Der durch zentrale Gesetzesänderungen verursachte enorme, dezentrale Anpassungsaufwand kann in der Vollzugsebene durch kooperative Arbeitsteilung deutlich reduziert werden.

Literaturverzeichnis

[EGovG] Bundesministerium des Innern: Gesetz zur Förderung der elektronischen Verwaltung sowie zur Änderung weiterer Vorschriften, 25. Juli 2013, Bundesgesetzblatt, Teil I Nr. 43, 2013.

[PWC14] pwc (Hrsg.): Prozessmanagement als notwendiger Baustein der Verwaltungsmodernisierung. http://www.pwc.de/de_DE/de/offentliche-unternehmen/assets/studie-prozessmanagement-register-2014.pdf (Zugriff: 14.04.2015), 2014.

[ITPR14] IT-Planungsrat: Föderales Informationsmanagement (FIM). http://www.it-planungsrat.de/DE/Projekte/Steuerungsprojekte/Steuerungsprojekte_NEGS/FIM/fim_node.html (Zugriff: 14.04.2015), 2014.

[PzMMV] Ministerium für Inneres und Sport Mecklenburg-Vorpommern: http://www.service.mv.de/cms/DLP_prod/DLP/Servicebereich/Kooperatives_E-Government/Prozessmanagement/Flyer_Projekt_PzM_MV.pdf, (Zugriff: 14.04.2015), 2013.

[Becker07] Becker, J.; Algermissen,L.; Pfeiffer, D.; Räckers, M.: Bausteinbasierte Modellierung von Prozesslandschaften mit der PICTURE-Methode am Beispiel der Universitätsverwaltung Münster. In: Wirtschaftsinformatik. 49 (2007) 4, S. 267–279, 2007.

[NFIM13] Bundesministerium des Innern, IT-Planungsrat, Ministerium der Finanzen Sachsen-Anhalt: Argumentationspapier Entscheidung und Begründung der Notationswahl für die Prozessmodellierung im FIM-Kontext, Berlin, Magdeburg, 2013.

[FIM13] Bundesministerium des Innern, IT-Planungsrat, Ministerium der Finanzen Sachsen-Anhalt: Feinkonzept FIM-Ordnungsrahmen. Entwurfsfassung vom 18.11.2013.

[DPMV] Ministerium für Inneres und Sport Mecklenburg-Vorpommern: Dienstleistungsportal. http://www.service.mv.de/cms/DLP_prod/DLP/Servicebereich/Kooperatives_E-Government/Prozessmanagement/index.jsp, (Zugriff: 15.04.2014).

[WIWI14] Häger, K.; Korzhynska, N.: Prozessmanagement in M-V und nationale Prozessbibliothek, 9. Wismarer Wirtschaftsinformatik-Tage (WIWITA), 12.-13.06.2014, Wismar, 2014.

[HoLa10] Hogrebe, F.; Lange, R.: Bausteine der Verwaltungsmodernisierung: Explorativer Vergleich von Methoden und Werkzeugen zur Visualisierung von Verwaltungsabläufen. In: Nüttgens, M.: Arbeitsberichte zur Wirtschaftsinformatik, 8/2010, Hamburg, 2010.

[eCH-0138] Schaffroth, C.D.M.: eCH-0138 Rahmenkonzept zur Beschreibung und Dokumentation von Aufgaben, Leistungen, Prozessen und Zugangsstrukturen der öffentlichen Verwaltung der Schweiz. 2012.

[AhWa14] Ahrend, N.; Walser, K.; Leopold, H.: Comparative Analysis of the Implementation of Business Process Management in Public Administration in Germany and Switzerland. In: 5th IFIP WG8.1 Working Conference on the Practice of Enterprise Modelling (PoEM 2012), November 7-8, 2012, Rostock, Germany

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Entwicklung eines generischen Gemeindereferenzprozessmodells in Anlehnung an das ARIS-Modell

Simon Enkerli, Konrad Walser[1]

Abstract: Die durchgängige Einführung von E-Government geht im Gegensatz zum Thema E-Business in der Privatwirtschaft eher beschaulich voran. Dies hängt u.a. auch damit zusammen, dass an der Schnittstelle von Verwaltung und IT noch größerer Entwicklungsbedarf ersichtlich ist. Konzeptionell ist dazu noch einige Grundlagenarbeit zu leisten, bis hier mehr Konsistenz in eine Verwaltungseinheits-übergreifende Entwicklung entsteht, sodass sich Interoperabilität zu einem durchgängig umgesetzten Prinzip etabliert. Dieser Beitrag schildert ein integriertes Konzept aus Sicht von Gemeindeverwaltungen, das aber auf allen föderalen Ebenen umgesetzt werden kann. Das Konzept dient der Abbildung von Geschäftsprozessen und den damit verbundenen Aspekten wie Formulare, Informationen, Leistungen, Aufgaben-, Kompetenz- und Veranwortungs-Zuordnungen sowie Organisations- und Governance-Aspekten. Als Grundlage dazu dient das für das vorliegende Vorhaben adaptierte ARIS-Modell. In einem iterativen Vorgehen ist das Modell mit Gemeindemitarbeitenden auf die Bedürfnisse einer Gemeinde angepasst worden. Die vorliegende Prozesslandkarte ist in die Bereiche Governance-, Führungs-, Kern- und Unterstützungsaufgaben gegliedert. Differenziert werden die folgenden Sichten auf Prozessebene: Eigentliche Prozesssicht, Organisationssicht mit Aufgaben, Verantwortungen und Zuständigkeiten, Informationssicht mit Formularen, Dokumenten und relevanten Gesetzen, Sicht auf Kontrollflüsse, Gateways und Geschäftsregeln sowie die Sicht auf die vom Prozessablauf betroffenen IT-Systeme.

Keywords: Prozessmanagement, Prozessreferenzmodell, Gemeinde, E-Government, Internes Kontrollsystem (IKS), ARIS-Modell.

1 Einleitung

1.1 Problemstellung

Die durchgängige Einführung von E-Government geht im Gegensatz zum Thema E-Business in der Privatwirtschaft in der Schweiz eher beschaulich voran ([GFS12]; [WR12]). Dies hängt u.a. auch damit zusammen, dass an der Schnittstelle von Verwaltung und IT größerer Entwicklungsbedarf besteht und konzeptionell einige Grundlagenarbeiten zu leisten sind, bis hier mit mehr Konsistenz eine Gemeindeübergreifende Entwicklung in Gang kommt, sodass sich Interoperabilität zu einem durchgängig umgesetzten Prinzip etabliert. Aktuell wird etwa in Deutschland, ausgehend vom Bundesministerium des Inneren und vom Ministerium der Finanzen in Sachsen, an einem sogenannten Föderalen Informationsmanagement (FIM) gearbeitet, mit dem löblichen Vorsatz, darin Prozess-,

[1] Berner Fachhochschule, E-Government-Institut, Brückenstrasse 73, CH-3005 Bern, simon.enkerli@e3ag.ch, konrad.walser@bfh.ch.

Formular- und Leistungsinformationen deutschlandweit zu standardisieren und auf Basis von (zentralen) Repositories zu integrieren ([AWR14]; [BMI12]). Das Projekt ist aktuell noch nicht abgeschlossen. Man darf gespannt sein, ob deutschlandweit eine Bereitschaft vorhanden ist, die erforderlichen Ressourcen und Infrastrukturen (Repositories) zur langfristigen Bereitstellung systematisch erfasster und abgelegter Informationen bereitzustellen und das Hosting zu gewährleisten. Denn, dazu wird eine Mittler- oder Broker-Infrastruktur zur Referenzierung der erforderlichen Informationen erforderlich sein. Eine wesentliche Aussage aus dem Projekt FIM ist, die erwähnten Informationen zusammenzubringen und gemeinsam bereitzustellen. Der „FIM-Baukasten" kann als etwas „zwischen den beteiligten Verwaltungseinheiten Liegendes" (als Mittlerinfrastruktur) verstanden werden, das allen Beteiligten im Austausch zur Verfügung stehen soll. Eine jedenfalls nicht explizit gemachte Intention kann die Standardisierung im deutschen E-Government sein. Standardisierung gewährleistet Interoperabilität. Hintergrund des Projektes war u.a. der Aufbau der deutschen Nationalen Prozessbibliothek, welche ein Bestandteil von FIM ist. Ausgehend von einem starken Föderalismus in Deutschland stellt sich die Kernfrage, über welchen Ansatz Standardisierung im E-Government erreicht werden kann. Bis heute ist dazu kein einheitlicher Ansatz ersichtlich. Die hier angesprochenen Mittler-Infrastrukturen sowie Vorgaben zur Gestaltung von Informationssystemen oder weitere Möglichkeiten sind im vorliegenden Beitrag nicht das Thema. Ähnliche Vorhaben wie das deutsche FIM-Vorhaben sind in der Schweiz als nicht realistisch verworfen worden. Es existierte hierzu auf Bundesebene ein Projekt „Reference eGov" [SEoJ], das sich neuerdings mehr in Richtung eines Behördenverzeichnisses entwickelt [PV14]. Analog zu Deutschland und dessen weiterem nationalen E-Government-Projekt Nationale Prozessbibliothek (NPB) existiert in der Schweiz ein priorisiertes Vorhaben von E-Government Schweiz, das Vorhaben eCH-Prozessaustauschplattform (B2.13) [Sc13], das in Teilen Anknüpfungspunkte zu FIM hat. Im Gegensatz zur hier diskutierten Problemstellung aus deutscher Sicht geht dieser Beitrag nicht von Mittler-Infrastrukturen als möglichem Handlungsfeld aus, sondern vielmehr von konzeptionellen Grundlagen, welche aus Sicht aller beteiligter Parteien, hier mit Schwerpunkt Gemeinden (aber auch Länder/Kantone), gegeben sein müssen, um im E-Government Interoperabilität zu erreichen. Dies kann in Form von Standards oder Implementierungen in Informationssystemen von Gemeinden erfolgen. Die große Anzahl von Gemeinden und die dadurch vorhandene Lösungsbreite von Leistungen, Prozessen, Formularen, Dokumentation(en)/-sformen (und zugrunde liegenden gesetzlichen Rahmenbedingungen), Modellen und das Informationsmanagement unterstützenden Informatikmitteln erschweren die elektronische Interoperabilität aus Sicht Gesetze, Organisationen, Prozesse, Semantik, etc. Würden bei Gemeinden oder anderen föderalen Verwaltungseinheiten überall einheitliche Prozessschritte, Formulare, etc. eingesetzt, würde dies (nicht abschließend) nicht nur die Kosten für die Pflege der Leistungen, den Dokumentationsaufwand und die Interoperabilitätskosten senken, sondern auch die elektronische Interaktion fördern, sowie die Transparenz für den Endkunden (Bürger, Kommune, Kanton/Land und Bund) erhöhen. Das vorliegende Konzept eignet sich so gesehen etwa für die Entwicklung eines entsprechenden Standards für die Strukturierung von Informationen bezüglich Prozessen, Formularen,

Leistungen, etc., und zwar nicht auf Basis einer Broker- oder Intermediärslösung, sondern als Konzept für die verwaltungsinterne Präsentation von Prozessen, Formularen, Informationen, etc. (Vgl. für das vorliegende Konzept [En14]). Durch die Entwicklung eines Gemeindereferenzprozessmodells haben die Gemeinden die Möglichkeit, sich einheitlicher zu strukturieren sowie die vorhandenen Prozesse und Erfahrungen untereinander auszutauschen. Zusätzlich vereinfacht ein solches Referenzprozessmodell die Transaktionen innerhalb der Behörden und standardisiert Abläufe sowie die dafür erforderlichen oder dadurch produzierten Informationen. In der Schweiz gab es in Richtung eines Domänenmodells zur Strukturierung des Verwaltungshandelns bereits erste Ansätze ([Wa13] in Anlehnung an [PTToJ]). Weniger weit gehend, respektive mehrheitlich auf die Vollzugsverwaltung fokussiert, existieren diverse eCH-Standards[2], welche aber noch nicht in Richtung eines holistischen Modells einer Gemeinde aus Sicht des Informationsmanagements gehen. Der hier vorliegende Ansatz zielt in Richtung eines holistischeren Modells.

1.2 Zielsetzung

Der vorliegende Beitrag verfolgt die folgenden Ziele: 1. Entwicklung eines (erweiterten) Geschäftsprozess-Referenzmodells für Gemeinden; 2. Zuordnung von in Zusammenhang mit dem Geschäftsprozess-Referenzmodell stehenden Informationen zu Formularen sowie Leistungen, die durch die Leistungen induzierten Prozesse, Rollen, Governance- und Entscheidungsaspekte sowie Anwendungsunterstützungen; 3. Bereitstellung eines systematischen Beitrags zur und Diskussion der integrierten Sicht der Gemeinde aus Sicht Informationsmanagement sowie zur Interoperabilität auf der kommunalen Ebene.

1.3 Methodisches Vorgehen

Dem vorliegenden Beitrag liegt ein iteratives Vorgehen im Rahmen einer qualitativen Forschung zugrunde. Das Forschungsvorgehen basiert auf [JP10] sowie auf explorativen und iterativ durchgeführten Interviews (vgl. [Fl07]) mit Gemeindemitarbeitenden. Die Interviews dienten der Eruierung der Ausrichtung des Prozessreferenzmodells, dem Bau eines Prototypen (wird aus Platzgründen in einem weiteren Beitrag thematisiert), der Überprüfung der Nutzungsakzeptanz des Prototypen, seiner Usability seitens der Gemeindemitarbeitenden sowie der Validierung (vgl. [HRP04]) des Modells über qualitative Interviews (Teile davon werden in einem Folgebeitrag präsentiert). Insofern haben

[2] Vgl. [DS12], [DOS12], [ML12], [BMR14]. Der Standard eCH-0138 mit dem Titel „Rahmenkonzept zur Beschreibung und Dokumentation von Aufgaben, Leistungen, Prozessen und Zugangsstrukturen der öffentlichen Verwaltung der Schweiz“ sowie der Standard eCH-0140 mit dem Titel „Vorgaben zur Beschreibung und Darstellung von Prozessen in der öffentlichen Verwaltung“ können hier spezifisch aufgeführt werden. Ein weiterer relevanter Standard lautet u.a. eCH-0158 „Modellierungskonventionen für die öffentliche Verwaltung“. In Teilen in Richtung eines holistischen Modells gehen E-Government-Architektur-Standards für die öffentliche Verwaltung (die eCH-Standards eCH-0122 bis eCH-0125); vgl. hierzu www.ech.ch. Die Konzentration dabei liegt indes auf der Vollzugsverwaltung. Die unterstützenden Prozesse der Verwaltung und die politischen Prozesse der Verwaltung werden – wenn überhaupt – nur vereinzelt oder am Rande darin abgehandelt.

aus wissenschaftstheoretischer Sicht im vorliegenden Beitrag sowohl Verhaltenswissenschaftliche als auch Design-Science-orientierte Forschungsansätze sich im iterativen Vorgehen abgelöst und zu den (in Kapitel 5) präsentierten Schlussfolgerungen geführt. In diesem Beitrag wird ein neu entwickeltes Gemeindegeschäftsprozess-Referenzmodell dargestellt. Wie erwähnt wird in einem weiteren Beitrag die empirische Untersuchung zur Nutzung und Usability des Gemeindegeschäftsprozess-Referenzmodells und dessen HTML-Prototypen präsentiert. Neben dem Design-Science-Research-Ansatz (Durchführung der sieben Schritte zur Erreichung der Rigorosität des Designs, dessen Artefakte, etc. bis hin zur hier vorliegenden Publikation) gelangte aus wissenschaftstheoretischer Sicht der Method-Engineering-Ansatz ([Ma95]; [VB03]) zur Anwendung. Die Konkretisierung des Gemeindeprozess-Referenzmodells erfolgte anhand der Adaptierung einer bestehenden Methode ([HRP04]; [Ma95]), hier des ARIS-Modells (nach [Sc92], [Sc96], [Sc01].

2 Theoretische Grundlagen

Im Folgenden ist auf verschiedene Gliederungsmöglichkeiten für Prozesse z.B. in Prozesslandkarten kurz einzugehen. Prozesslandkarten können etwa wie folgt gegliedert werden: Management-, Kern- und Supportprozesse (vgl. hierzu ausgehend von eCH-Standards die Abbildung 1).

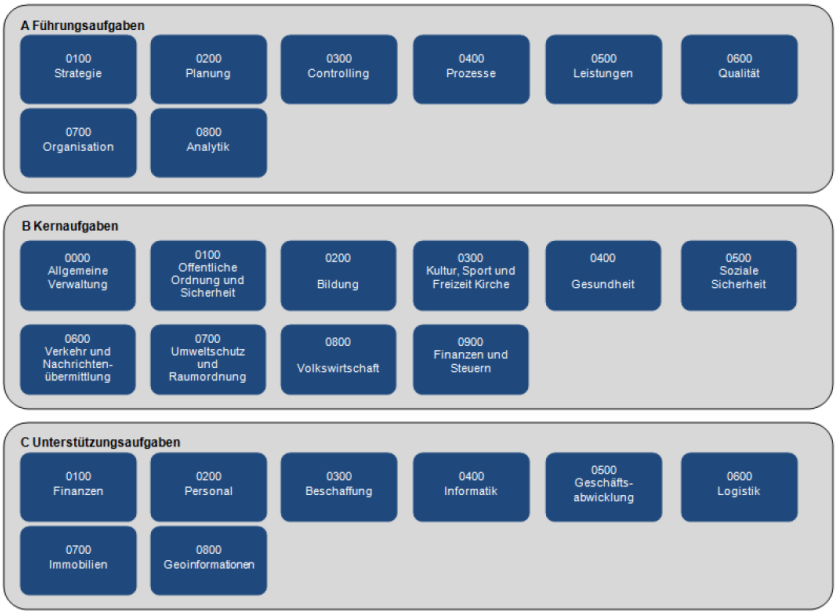


Abbildung 1: Strukturierung der Vollzugsaufgaben einer Gemeinde nach Aufgabentypen und -feldern (in Anlehnung an eCH-0145).

Für die Verwaltung ist in der Schweiz ein erster Versuch einer entsprechenden Darstellung vom Verein eCH ([Sc10]; E-Government-Referenzmodell Gemeinden) sowie in eine holistischere Richtung durch [Wa13] (Domänenmodell der Verwaltungsorganisation, teilweise in Anlehnung an [PTToJ]) unternommen worden. Der Verein eCH [Sc10] teilt die einzelnen Gemeindeprozesse in die drei Ebenen „Management, Aufgaben und interne Querschnittsaufgaben" auf. Der entsprechende Standard ist nie fertiggestellt worden. [Wa13] benennt die drei Ebenen „Politische Prozessdomäne", „Leistungsverwaltungs-Prozessdomäne" und „Prozessdomäne für unterstützende Prozesse" sowie „strategische Prozessdomäne" (je separate strategische Führung der vorgenannten Domänen), für die weiter Prozess(bereich)e differenziert werden können. Weiter wurde durch [LZR05], basierend auf dem St. Galler Managementmodell, ein Prozessmodell propagiert, das sich zur Darstellung von Geschäftsprozessen in der öffentlichen Verwaltung eignet. Eine spezifische Eigenart dieses Modells ist, dass es die Umgebung und die diversen daraus resultierenden Einflussfaktoren auf die Ausprägung des Geschäftsprozessreferenzmodells eines Unternehmens zeigt. Das St. Galler Modell teilt die Prozesse ebenfalls in die drei Kategorien Management-, Kern- und Supportprozesse ein [RS05]. Der eCH-Standard eCH-138 ([DS12]; [DOS12]) zeigt die gleiche Aufteilung der Prozesse in Management-, Kern- und Unterstützungsaufgaben. In diesem Modell werden die Kernaufgaben noch in zwei unterschiedliche Blöcke unterteilt, nämlich in „Aufgaben mit Vollzugscharakter" und „Aufgaben mit Voraussetzungscharakter" ([DS12]; [DOS12]). Als weiteres kategorisierendes Referenzmodell kann etwa das hinlänglich bekannte Wertschöpfungsmodell nach Porter (Vgl. [Po92]) erwähnt werden (Differenzierung in Kernprozesse und Unterstützungsprozesse). Auch die eCH-Standards eCH-0145 (Aufgabenlandkarte; vgl. www.ech.ch) sowie eCH-0122 (Architektur E-Government Schweiz [ML14]) können als Referenzprozessmodelle verstanden werden. In Deutschland existiert ebenfalls ein entsprechendes Referenzmodell für Gemeinde-Verwaltungsprozesse [BV12]. Darin wird ersichtlich, dass Deutschland mit der Strukturierung kommunaler Prozesse ähnliche Wege geht wie die Schweiz. Der im vorliegenden Beitrag geschilderte Ansatz konzentriert sich auf die Vollzugsprozesse einer Gemeinde (Leistungsverwaltung) und seiner Exekutive und nicht auf Prozesse der politischen Verwaltung oder des Verwaltungs-Supports. Der vorliegende Ansatz thematisiert ebenfalls nicht: Legislative oder judikative Prozesse und entsprechende Prozesslandkarten. Diese wären separat zu behandeln. Sie können aber vom entwickelten Referenzmodell profitieren. Der vorliegende Beitrag konzentriert sich auf die fachliche und nicht auf die technische Seite des Prozessmanagements (Vgl. [FR12]). Er stellt jedoch durch den (konzeptionell) entwickelten Prototypen technische Umsetzungsvoraussetzungen mit Anknüpfungspunkten zur (technischen) Interoperabilitätsperspektive (Interfaces) zur Verfügung. Ein frühes Modell, das für das erweiterte Management von Geschäftsprozessen eine zentrale Rolle spielte, war dasjenige von Scheer (Vgl. [Sc92]; [Sc96]; [Sc01] ARIS). Dieses präsentiert im Zusammenhang mit Geschäftsprozessen die fünf folgenden Sichten auf Informationssysteme und zeigt deren Integration auf: Funktionssicht, Organisationssicht, Datensicht, Leistungssicht sowie Steuerungssicht. Jede der entsprechenden Sichten ist in drei Beschreibungsebenen unterteilt: Fachkonzept, Datenverarbeitungskonzept und Implementierungskonzept.

Somit kann das ARIS-Modell als frühe Grundlage eines Unternehmensarchitektur-Managementwerkzeugs (EAM) verstanden werden. Das ARIS-Modell hat im deutschsprachigen Umfeld eine gewisse Bekanntheit erreicht, ist über die Jahre hinweg aber nur mehr eines unter vielen EAM-Werkzeugen, die heute zur Verfügung stehen [Ma11]. Der Aufbau von ARIS macht ausgehend von seiner Eingängigkeit her für die Anwendung im Gemeindereferenzmodell zielsetzungsmäßig Sinn, weil dadurch genau diese basalen Zusammenhänge innerhalb der Gemeinde anhand einer Adaption des ARIS-Modells veranschaulicht werden können.

Gemeinden können als kleinste granulare Einheit des Staatsaufbaus verstanden werden. Sie leisten vor Ort die Bearbeitung von Bürger- und Unternehmensanliegen ausgehend von einem gesetzlichen Auftrag. Über den Gemeinden sind üblicherweise Staatsgebilde oder -strukturen aufgebaut, welche die Gemeinden zu Staaten zusammenfassen. Die Staaten werden üblicherweise auf einer nationalen Ebene zu einem Staatenbund/Bundesstaat zusammengefasst (föderaler Staatsaufbau, Subsidiarität als Kernprinzip der Aufgabenteilungen zwischen den verschiedenen föderalen Ebenen [Li05]). Ausgehend vom föderalen subsidiären Staatsaufbau können nun Aufgaben definiert werden, welche von den Gemeinden übernommen und ausgeführt werden. In der Schweiz ist hierfür der sogenannte E-Government-Standard eCH-0145 entwickelt worden [DOS13]. Dieser Standard sieht die Differenzierung der Vollzugsaufgaben der Gemeinde in der Form vor, wie sie Abbildung 1 zeigt. In eCH-0145 werden die Aufgaben nicht nur bezeichnet, sondern auch nummeriert. Die Nummerierung kann entsprechend des Konkretisierungsgrads oder der Differenzierung über die Nummerierung im Dezimalklassifikationssystem verfeinert werden, was der Orientierung im Gesamtmodell dienlich ist. Da die Nummerierung aktuell ausgehend vom Standard eCH-0145 nicht ganz konsistent ist (z.B. kann die Nr. 100 auf allen Prozessebenen vorkommen), werden die Prozesse im vorliegenden Beitrag, falls Erwähnung findend, hierarchisch mit A, B und C für Führungs-, Ausführungs- und Unterstützungsprozesse gekennzeichnet. Im weiteren Verlauf des Beitrags erfolgt die Fokussierung auf das ARIS-Modell von Scheer [Sc92] und die eigene Adaptierung für Gemeindebedürfnisse. Die Evaluation ist durch ein halbes Dutzend Gemeindemitarbeitende erfolgt. Alle erwähnten Modelle wurden in die Evaluation miteinbezogen. Resultat: Das ARIS-Modell eignet sich für die vorliegenden Ziele am besten.

3 Modellentwicklung als Prozess basierend auf Interviews

Im Folgenden wird auf Basis von fünf durchgeführten Iterationen die Entwicklung des Gemeinde-Referenzprozessmodells präsentiert (Übersicht im Anhang dieses Beitrags).

3.1 Entwicklungszustand 1 – Entwicklung der Landkarte auf Ebene 0

Dieser Entwicklungszustand geht von den bereits geschilderten Modellen zur Prozessdarstellung und deren Einbettung wie in Kapitel 2 (Theoretische Grundlagen) präsentiert aus. Daraus erfolgt die Ableitung einer Prozesslandkarte für die Vollzugsaufgaben und -

prozesse der Gemeinde. Zur Entwicklung trugen Experteninterviews entsprechend der folgenden Frage bei: „Welches Modell favorisieren sie für die Prozessdarstellung“?. Das Resultat lautete: eCH-0145 als Favorit für die Übersichtsdarstellung der Vollzugsaufgaben der Gemeinde. Der Grund dafür ist, dass eCH-0145 eine große Nähe zu HRM2 hat (Vgl. zu HRM und HRM2, ausgeschrieben Harmonisiertes Rechnungsmodell, den folgenden Link: hrm2.zh.ch. Das HRM-Modell zeigt die an herkömmliche Buchführungsmethoden angenäherte Entwicklung der Buchführung der öffentlichen Verwaltung auf). Das Resultat ergibt die folgende Prozesslandkarte aus Gemeindesicht für die Vollzugsaufgaben.

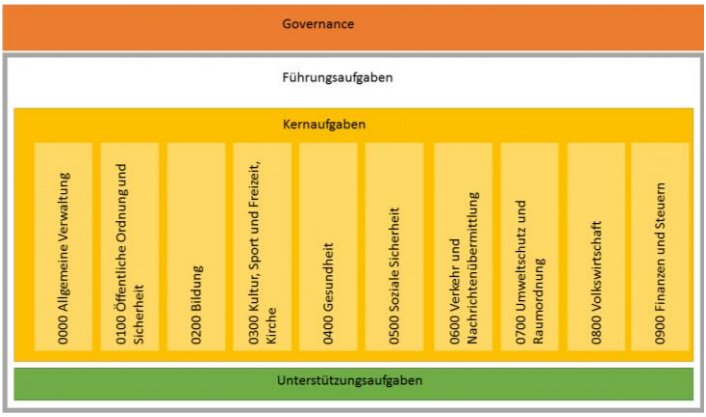


Abbildung 2: Prozesslandkarte Ebene 0.

3.2 Entwicklungszustand 2 – Entwicklung der Bereichsauswahl auf Ebene 1

Hier erfolgt die Ergänzung der Prozesslandkarte um die nächsttiefere Ebene 1. Dies führt zu folgenden Anpassungen: Die Übernahme des eCH-Standards eCH-0145 genießt unter den Interviewten Sympathien, da dieser schweizweit im Einsatz ist und für die gesamte Schweiz Gültigkeit hat. Dies führt zu einer weitgehenden Übernahme der in eCH-0145 konkretisierten Feindifferenzierungen der Aufgaben.



Abbildung 3: Bereich Governance und dessen Prozesse auf der Ebene 1.

Im vorliegenden Fall besteht eine Auswahlmöglichkeit bezüglich: Führungsaufgaben, Kernaufgaben oder Unterstützungsaufgaben. Einer dieser Bereiche ist zu wählen um

darunter die Konkretisierung der möglichen Teilaufgaben und -prozesse einzusehen. Für die Governance-Ebene resultierten aus den Interviews analoge Aufgabengebiete wie diese in COBIT für den Governance-Teil vorgeschlagen werden [ISA12]. Pflege und Transparenz, Risiko- und Ressourcenmanagement sowie Wertbeitrag. Dies bedeutet konkret je genanntem Bereich das Folgende: 1. Pflege: Einführung eines Pflegeprozesses für das Gemeinde-Referenzprozessmodell; 2. Transparenz: Prozess zur Sicherstellung und Überprüfung, dass das vorhandene implizite Wissen der Mitarbeitenden vollständig in den Prozessen und den dazugehörigen Hilfsmitteln, wie z.B. Dokumente oder Formulare, externalisiert ist. 3. Risikomanagement: Integration Gemeindeverwaltungs-externer und -interner (gesetzlicher und sonstiger) Anforderungen um das Risiko zu minimieren, an den Zielgruppen vorbei Dienstleistungen bereitzustellen. Von besonderer Bedeutung ist dieses Thema etwa bei finanziellen Transaktionen. Hier wird typischerweise von internen Kontrollsystemen (IKS) gesprochen, welche als Instrument des Risikomanagements eingeführt werden soll(t)en (hierzu unternimmt eCH aktuell Anstrengungen zur Generierung neuer eCH-IKS-Standards). 4. Ressourcenmanagement: U.a. aus Personalsicht laufende Überprüfung der Ressourcensituation in den einzelnen Verwaltungsbereichen der Gemeinde, um sicherzustellen, dass Ressourcen adäquat und gemäß gesetzlichem Auftrag eingesetzt werden. 5. Wertbeitrag: Hinterfragung der Dienstleistung dahingehend, dass dem Bürger oder dem Unternehmen nur das angeboten wird, was er/es (aus gesetzlicher oder nicht gesetzlicher Sicht) wirklich braucht. Dies ergibt für den Bereich Governance in der Prozessübersicht die in Abbildung 3 dargestellte Teilübersicht über die Governance-Prozesse der Gemeinde.

3.3 Entwicklungszustand 3 – Entwicklung der Prozessauswahl auf Ebene 2

Im vorliegenden Zustand 3 geht es um die Konkretisierung der Prozessauswahl, in welcher eine Nähe zum eCH-Standard eCH-0145 gemäß Interviews als gesetzt gilt.

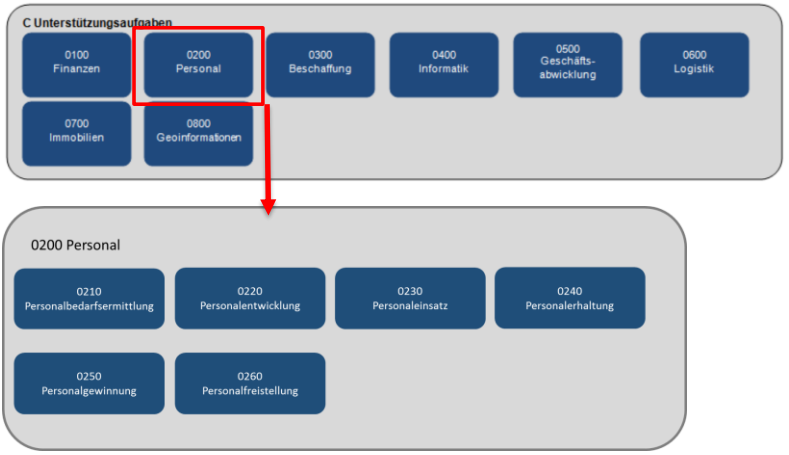


Abbildung 4: Prozessbereich Personal und dessen Teilaufgaben.

Der Zweck der Ebene 2 lautet wie folgt: Gewährleistung der Prozessübersicht eines Bereichs. Aus Sicht des Nutzers des Referenzprozessmodells kann dieser ausgehend von der vorhergehenden Ebene 1 auf einen Bereich klicken und erhält dann die Übersicht über die zugehörigen Prozesse auf Ebene 2. In der Abbildung 4 wird dies anhand der Unterstützungsaufgaben dargestellt und darin des Aufgabenbereichs Personal. Im Bereich etwa der Personalprozesse (und der Supportprozesse überhaupt) wird von eCH keine weitere Hilfestellung bezüglich erforderlicher Prozesse angeboten. Im vorliegenden Fall wird dazu auf [ZBT01] referenziert.

3.4 Entwicklungszustand 4 – Entwicklung der Prozessdarstellung auf Ebene 3

Zusammenfassend zeigten sich die Interviewpartner zufrieden mit dem Ansatz der verschiedenen Sichten. Aus den Interviews kann auch die Schlussfolgerung gezogen werden, dass die Sichten ein integriertes Bild des Prozessmanagements in der Gemeinde ergeben. Hilfestellungen im Bereich der Modellierung der Prozesse bieten die weiter oben referenzierten Standards, u.a. die Modellierungskonventionen von eCH zum Standard BPMN 2.0. Aus Sicht eCH ist bei den übrigen Sichten darauf hinzuweisen, dass hier möglicherweise weitere Standards erstellt werden können, beispielsweise für Kontrollflüsse, IKS-Anforderungen (in Arbeit) in Relation zu den Kontrollflüssen, Organigrammen und damit verbunden die Konkretisierung von Verantwortlichkeiten und Zuständigkeiten, Informationsbereitstellungen und Beratungsmiteinbezüge in Prozessen. Ebenfalls macht es Sinn, Prozessdokumentationen und etwa auch Dokumentationen zu Formularen zu standardisieren. Auf der Ebene 3 finden überdies die Integrationen der Sichten Organisation, Prozesse, Funktionen, Kontrollflüsse sowie der Beschreibung von Entscheidungspunkten in Prozessen sowie Geschäftsregeln für die jeweiligen Prozesse statt.

3.5 Entwicklungszustand 5 – Adaptierung der Prozessdarstellung 4

Im Zustand 5 geht es um Verfeinerungen des Entwicklungszustands 4. Aus den Interviews resultierte, dass jeder Prozess(schritt) mit einem RACI-Chart versehen werden soll, was entsprechend nachvollzogen wird. Ebenfalls wird in Zustand 5 grafisch dokumentiert, in welchen Informationssystemen der Prozess implementiert wird. Dabei wird zwischen den Anwendungstypenkategorien wie folgt unterschieden (in Anlehnung an [WRoJ]): Front Office Applikation (FO), Back Office Applikation (BO), Schnittstellen (IF), Services (SV). Ein Beispiel einer entsprechenden Darstellung wird in Abbildung 5 präsentiert.

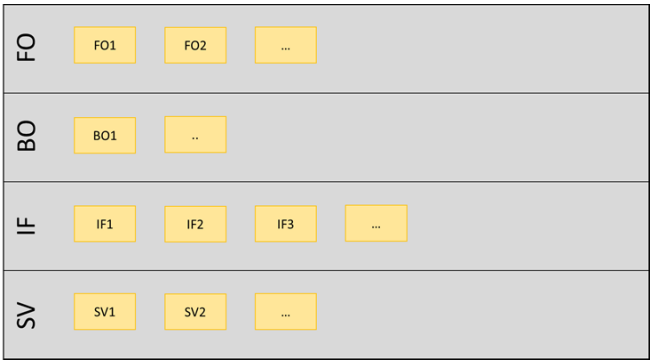


Abbildung 5: Fachanwendungskategorien im Gemeindeumfeld, über welche Gemeindegeschäftsprozesse instanziiert werden können.

3.6 Das Gemeindeprozess-Referenzmodell im Überblick

Das oben kurz beschriebene Modell ARIS-House nach [Sc92] wurde ausgehend von den Interviews mit Gemeindemitarbeitenden als Referenzgrundlage für das vorliegende Gemeindeprozess-Referenzmodell verwendet [Sc9], [Sc01], [SJW05]. Darauf bauen die einzelnen Weiterentwicklungen auf, die vom Vorgehen her an [VB03] angelehnt sind. Die Modellentwicklung erfolgte wie dargestellt in 5 Iterationszyklen, welche sich von der Ebene 0 bis auf die Ebene 4 des Gemeindereferenzprozessmodells erstreckten. Aus Platzgründen können diese Zustände hier nicht detaillierter beschrieben werden (Einsichtnahme kann bei den Autoren des Beitrags anbegehrt werden). Der Endzustand des Modells (auf Ebene 4) präsentiert sich wie folgt: 1. Gemeindeorganigramm mit detaillierten Darstellungen der Zuständigkeiten in Form von RACI-Charts (wer ist verantwortlich (A), wer zuständig für die Ausführung (R), wer zu informieren (I), wer zu konsultieren (C). In der entwickelten Darstellung für Gemeinden werden nur R und A angewandt). 2. Funktionsdiagramm mit Aktivitäten, deren Beschreibung und Erweiterung um die Zuständigkeiten analog RACI-Charts (hier nun alle vier Bereiche umfassend zu modellieren). 3. Kontrollfluss-Diagramm, basierend auf den funktionalen Teilschritten wie unter Funktionen aufgeführt. 4. Gateway-Diagramm, zur differenzierten Darstellung der Entscheidungspunkte im Prozess. 5. Geschäftsregeln mit Darstellung der (Vorgangs-)Regeln im Prozess, welche Gültigkeit haben. 6. Datendiagramm; hier nicht als Entity Relationship Diagramm interpretiert, sondern als Informations(erfassungs)grundlagen zum Prozess, umfassend Formulare, Dokumente – etwa in Form von Checklisten, Gesetzen, Verordnungen, welche Gesetze in deren Umsetzung konkretisieren. 7. IT-Systemsicht: In dieser Sicht werden die IT-Systeme charakterisiert, über welche die Prozessabwicklung erfolgt. Dies können Schnittstellen, Front- und Back-Office-Systeme oder auch Bussysteme für die Integration der verschiedenen Anwendungen sein. Dies führt zum im Anhang präsentierten finalen Modell, über welches die weiteren Ausführungen dazu konkretisiert werden können. Ausgehend von der vorliegenden Präsentation eines Gemeindeprozess-Referenzmodells lässt sich vereinfacht erahnen, dass man über

die Landkarte (Ebene 0) systematisch über die Ebenen 0 bis 4 feiner differenzieren kann (Drill-down) und dass Navigationsmöglichkeiten für Gemeindemitarbeiter vorhanden sind. Dies wurde in Form eines Prototyps konkretisiert, der für einige wenige ausgewählte Prozesse diesen Drill-down beinhaltet (Darstellung in einem weiteren Beitrag). Die Idee ist, diese prototypische Darstellung künftig für alle Prozesse einer Gemeinde zu vervollständigen, falls weitere Sichten unterschiedlicher Ebenen einzuschließen sind.

Ebene	Grundlage	Beschreibung
0	Geschäftsprozess-Landkarte verknüpft mit dem Standard eCH-0145	Die erste Ebene zeigt die Landkarte auf. Sie bietet die drei vom Standard eCH-0145 definierten Bereiche: Führungsaufgaben, Kernaufgaben und Unterstützungsaufgaben. Hinzugefügt wurde ein Governance-Bereich.
1	Aufgegliederter Standard eCH -0145	Die zweite Ebene zeigt die Prozessgruppen der ausgewählten Bereiche auf, z.B. alle Prozessgruppen des Bereichs Kernaufgaben.
2	Aufgegliederter Standard eCH-0145	Die dritte Ebene zeigt die Prozessgruppen zum ausgewählten Bereich auf. Z.B alle Prozessgruppen des Bereichs allgemeine Verwaltung.
3	Prozessdarstellung	Die vierte Ebene zeigt entweder bereits die Prozessdarstellung in den Sichten Organisation, Daten, Funktion, IT-System und Steuerung oder es werden weitere Prozessgruppen gebildet.
4	Prozessdarstellung	Die fünfte Ebene ist mit konkretisierten Beispielprozessen die letzte Ebene im vorliegenden Beitrag, welche als Prozessdarstellung in die Sichten Organisation, Daten, Funktion, IT-System und Steuerung aufgeteilt ist.

Abbildung 6: Ebenen des Gemeindeprozess-Referenzmodells mit Charakterisierung der Sichten und Ebenen.

Die Realisierung erfolgt unterschiedlich, sei es ausgehend von Prozessmodellierungsanwendungen oder als Webapplikation z.T. basierend auf einem Prozessmodellierungstool. Möglicherweise umfassen einzelne Prozessmodellierungswerkzeuge entsprechende Darstellungsmöglichkeiten. Dies wurde ausgehend vom vorliegenden Beitrag nicht untersucht und wäre eine eigene Untersuchung wert. Die entsprechende Abbildung im Anhang zeigt basierend auf den verschiedenen Ebenen eine Übersichtsdarstellung über das gesamte Modell. Basierend darauf kann von oben nach unten auch das Vorgehen des Drill-downs von der Ebene 0 auf die Ebene 4 dargestellt werden.

4 Nutzen, Nutzungsmöglichkeiten und Erweiterungsmöglichkeiten des Gemeinde-Prozessreferenzmodells

Die folgenden Nutzen werden mit dem vorliegenden Geschäftsprozess-Referenzmodell

für Gemeinden ausgehend von der obigen Darstellung verbunden: Es ist eine einfache Navigation in einem Referenzprozessmodell der Gemeinde basierend auf einem Mehrebenen-Modell und basierend auf Ebenen-spezifischen Informationsbereitstellungen gegeben. Es ist eine logische Ordnung der entsprechenden Informationen aufgrund einer Nummerierung und Kategorisierung der Information gegeben. Es ist eine problemlose Navigation in die Bereiche, in denen ein Mitarbeitender typischerweise tätig ist, gegeben. Über die reine Prozessorientierung hinausgehend sind die folgenden weiteren Artefakte darin positionierbar: Funktionsdiagramme, Entscheidungsrechte der Beteiligten, IKS-relevante Informationen etwa zu Risiken. Es ist eine einfache Top-down- und Bottom-up-Navigation möglich. Es ist eine Suche über Prozesse und Informationen möglich. Die folgenden Nutzungs- und Erweiterungsmöglichkeiten für das vorliegende Geschäftsprozess-Referenzmodell für Gemeinden sind denkbar: Das Geschäftsprozessreferenzmodell kann als Navigationsinstrument in der Gemeindeverwaltung dienen. Es stellt einen expliziten „Speicher von Wissen“ für alle Mitarbeitenden im Gemeindeumfeld dar. Es stellt die Grundlage für die Einführung (und die Führung) von Mitarbeitenden in die Prozesse (sowie deren Führung in) der Gemeinde dar. Ebenso mag es die Grundlagendokumentation dafür bieten, was Mitarbeitende können, wissen und lernen müssen. Es stellt die Grundlage für die Definition und Kategorisierung von Kennzahlenrastern und Auswertungsmöglichkeiten zur Führung der Gemeindeverwaltung dar, aber auch zur Bildung einer Struktur z.B. in Richtung eines Front-Office-Informationsspeichers, d.h. zur Information, Beauskunftung, Beratung sowie Nachsorge in einem generischen Front Office einer Gemeinde, sei es über Web, Schalter, Telefon oder mobile Kanäle, etc. Es stellt die Grundlage zur Bildung von Anforderungskatalogen oder Stellenbeschrieben einer Gemeinde zur Beschaffung/Weiterbildung von Personal sowie zur Entwicklung neuer Software dar. Es stellt die Grundlage zur Weiterentwicklung und Integration von Informationssystemen durch die darunter liegende Präsentation eines Informationsmodells der Gemeinde dar. Es stellt die Bildung von Repositories zur standardisierten Entwicklung von E-Gov-Services bereit. Es stellt die Grundlage für weitere Standardisierungen im Gemeindekontext im weiteren oder näheren Umfeld des Geschäftsprozessmanagements bereit. „Last but not least“ dient es möglicherweise auch der Spezifikation von Meldungstypen, welche beispielweise in der Gemeinde oder über die Gemeinde hinaus ausgetauscht werden können müssen und es dient entsprechend der strukturierten Konkretisierung grundsätzlich der Interoperabilität.

5 Zusammenfassung und Ausblick

Der vorliegende Beitrag schildert ausgehend von einem qualitativen und iterativen Forschungsvorgehen die Entwicklung eines Geschäftsprozess-Referenzmodells basierend auf bereits bestehenden Artefakten und Methodologien, die aber für Gemeinden in diesem Beitrag neu adaptiert werden. Das Ziel des Beitrags ist es, aus Gemeindesicht zu einem Referenzprozessmodell zu kommen, das im Gegensatz zu einer einfachen Sammlung und Dokumentation von Prozessen über deren Ziel weit hinausgeht. Die Notwendigkeit eines entsprechenden Modells scheint unbestritten. Der Vorteil eines entspre-

chenden Modells liegt darin, dass es von der Implementierung her abstrahiert und daher entsprechenden Mittler-Modellen (von Staaten oder Gemeinden bereitzustellende Repositories, Broker, etc.) wie im Falle des Föderalen Informationsmanagements FIM (Perspektive von außen nach innen) klar überlegen ist (keine umfangreichen Mittlerinfrastrukturen erforderlich), indes die Mehrheit der dort thematisierten Elemente integriert. Ein entsprechendes Mittlermodell analog zu FIM erfordert einen zentralen Aufbau, Betrieb und Wartung von Repositories. Dies scheint in sehr föderalen Systemen wie in vielen europäischen Ländern nicht möglich oder zielführend (zentrales Problem Föderalismus), weil vielfach kein Konsens darüber gefunden werden kann, wie dieses finanziert werden soll. Vielmehr bietet das vorliegende Geschäftsprozess-Referenzmodell für Gemeinden Grundlagen für die Weiterentwicklung von Gemeindeinformationssystemen und entsprechenden Anwendungs-Landschaften (von innen nach außen) etwa durch Integratoren (Unternehmensarchitektur der Gemeinde) in diversen Richtungen und Ausprägungen. Zudem bietet das vorliegende Referenzprozessmodell Möglichkeiten, etwa auch systematisch Schnittstellen und Datenaustausche generisch zu differenzieren und entsprechend zu standardisieren (für Meldedaten ist das beispielsweise im Einwohnerwesen ist das bei eCH bereits schon erfolgt). Das Refenzprozessmodell der Gemeinde basiert auf folgenden Sichten: Organisation, Funktion, Daten und einer alle drei Dimensionen verbindenden Steuerungsebene. Im Grunde wird dieses Set an Perspektiven ergänzt um RACI-Charts aus Prozesssicht, im Scheer Modell noch mit Funktionsdiagrammen umschrieben, Kontrollfluss-Grafiken, u.a. neben dem Prozessmodell für IKS-Zwecke verwendbar, sowie Architekturgrafiken dazu, über welche Informationssysteme und deren Komponenten Geschäftsprozesse ablaufen und was für Schnittstellen und Services hierzu zur Verfügung zu stellen sind. Insofern ist über das vorliegende Geschäftsprozessreferenzmodell für die Gemeinde auch Anschlussfähigkeit für die technische Umsetzung von Geschäftsprozessen in Gemeinden gegeben. Das Modell ist in vielerlei Richtungen ausbaubar und genießt in seiner heutigen Form den Charme, dass es noch einigermaßen überschaubar und nicht bereits zu komplex ist.

Literaturverzeichnis

[AWR14] Ahrend, N.; Wolf, P.; Räckers, M.; Dentschev, M.; Jurisch, M. (2014): Federal Information Management – Context and Effects. In: Ploedereder, E.; Grunske, L.; Schneider, E., Ull, D. (Hrsg.): Proceedings der INFORMATIK 2014 - Big Data – Komplexität meistern, vom 22. – 26. September 2014 in Stuttgart, GI Lecture Notes in Informatics, S. 1499-1511.

[BMI12] Bundesministerium des Inneren (2012): FIM – E-Government mit Zukunft - Die einheitliche Basis für eine effiziente Verwaltung in Kommunen, Ländern und Bund. BMI, Berlin/Bonn.

[BMR14] Bosshart, E.; Märki, M.; Rigert, B.; Spöcker, N.; Tanner, C. (2014): eCH-0158 – BPMN-Modellierungskonventionen für die öffentliche Verwaltung: Auf: http://www. ech.ch/alfresco/guestDownload/attach/workspace/SpacesStore/3140e2f9-74c8-4ebe-aef8-0f7732061e0f/STAN_d_DEF_2014-09-04_eCH-0158_V1.1_BPMN-Modellie rungs Modellierungskonventionen%20für%20die%20öffentliche%20Verwaltung.pdf (Aufruf per 2014-11-03).

[BV12] Bundesverwaltungsamt (2012): Leitfaden für die Erhebung von Geschäftsprozessen im Bundesministerium des Innern und seinen nachgeordneten Behörden, Köln: Bundesverwaltungsamt.

[DOS12] Desobry, C.; Opitz, P.; Schaffroth, M.; Schneider, S. (2012): eCH-0140 – Vorgaben zur Beschreibung und Darstellung von Prozessen der öffentlichen Verwaltung der Schweiz. Auf: http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0140&documentVersion=1.00 [Zugriff am 03 12 2013].

[DOS13] Delafontaine, S.; Opitz, P.; Schaffroth, M. (2013): Verein eCH - eCH-0145 Aufgabenlandkarte der öffentlichen Verwaltung der Schweiz. Auf: http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0145&documentVersion=1.0 [Zugriff am 02 02 2014].

[DS12] Desobry, C.; Schaffroth, M. (2012): eCH-0138 – Rahmenkonzept zur Beschreibung und Dokumentation von Aufgaben, Leistungen, Prozessen und Zugangsstrukturen der öffentlichen Verwaltung der Schweiz. Auf: http//www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0138&documentVersion=1.00 [Zugriff am 03 12 2013].

[En14] Enkerli, S. (2014): Gemeinde-Referenzprozessmodell – Ausarbeitung einer generischen Gemeindelandkarte mit Modellierung ausgewählter Elementarprozesse. Master Thesis in Wirtschaftsinformatik, Berner Fachhochschule, E-Government-Institut, Bern.

[Fl07] Flick, U. (2007): Qualitative Sozialforschung. Eine Einführung. Reinbek bei Hamburg: Rowohlt.

[FR12] Freund, J.; Rücker, B. (2012): Praxishandbuch BPMN 2.0. München: Carl Hanser Verlag GmbH & Co. KG.

[GFS12] GFS.Bern (2012): E-Government in Kantonen auf dem Vormarsch, aber durch fehlende Personalressourcen gebremst. Schlussbericht Verwaltung und E-Government 2012, Issue GFS.Bern.

[ISA12] ISACA (2012): COBIT 5 – Rahmenwerk für Governance und Management der Unternehmens-IT. Rolling Meadows, IL 60008, USA: ISACA.

[JP10] Jonker, J.; Pennink, B. (2010): The Essence of Research Methodology. Heidelberg: Springer.

[Li05] Linder, W. (2005): Schweizerische Demokratie – Institutionen, Prozesse, Perspektiven, Haupt, Bern.

[LZR05] Landolt, P.; Zuppiger Ritter, I. (2005): Strategisches Management in der öffentlichen Verwaltung, St. Gallen: s.n.

[Ma11] Matthes, D. (2011): Enterprise Architecture Frameworks Kompendium: Über 50 Rahmenwerke für das IT-Management. Springer Science+Business Media, Berlin.

[Ma95] Mayer, R.J. (1995): Information Integration for Concurrent Engineering (IICE) – Compendium of Methods Report. Ohio, Human Resource Directorate Logisitics Reseearch Division.

[MW14] Müller, W.; Lindner, H. (2014): eCH-0122 – Architektur E-Government Schweiz: Grundlagen: Auf: http://www.ech.ch/alfresco/guestDownload/attach/workspace/SpacesStore/e9558674-3fa8-4fb4-9c0c-399b62672f59/BEST_d_DEF _2014-06-12_eCH-0122_V1-0_eGovCH-Grundlagen.pdf (Aufruf per 2014-11-03).

[Po92] Porter, M. E. (1992): Wettbewerbsvorteile (Competitive Advantage): Spitzenleistungen erreichen und behaupten. 3.Auflage Hrsg. Frankfurt am Main: Campus.

[PTToJ] Peristeras, V.; Tsekos, T.; Tarabanis, K. (o.J.): E-Government or E-Governance? Building a Domain Model for the Governance System, United Nations Thessaloniki Centre, University of Macedonia: s.n.

[PV14] Priorisiertes Vorhaben B2.13 (2014). Auf: http://www.refegov.ch/de/aktuell /aktionsprogramm-egov/index.html (Aufruf per 2014-10-23).

[RS05] Rüegg-Stürm, J., 2005. Das neue St. Galler Management-Modell. s.l.:Palgrave Macmillan.

[Sc01] Scheer, A.-W. (2001): ARIS - Modellierungsmethoden, Metamodelle, Anwendungen. Vierte Auflage Hrsg. Heidelberg: Springer-Verlag Berlin.

[Sc10] Schärli, T. (2010): E-Government Referenzmodell Gemeinden. Zürich: s.n.

[Sc13] Schaffroth, M. (2013): B1.13 – eCH-Prozessaustauschplattform für Gemeinden und Kantone. Auf: http://www.ech.ch/alfresco/guestDownload/attach/workspace/SpacesStore/06db5eb7-3691-47ed-b78f-f12eb466aa10/eGov_2-2013__41-41_Schaffroth.pdf (Aufruf per 2014-10-23).

[Sc92] Scheer, A.-W. (1992): Architektur integrierter Informationssysteme. Springer, Berlin.

[Sc96] Scheer, A.-W. (1996): ARIS-House of Business Engineering: Von der Geschäftsprozeßmodellierung zur Workflow-gesteuerten Anwendung; vom Business Process Reengineering zum Continuous Process Improvement.. Saarbrücken: Institut für Wirtschaftsinformatik (IWi) an der Universität des Saarlandes.

[SJW05] Scheer, A.-W.; Jost, W.; Wagner, K. (2005): Von Prozessmodellen zu lauffähigen Anwendungen. Heidelberg: Springer-Verlag Berlin.

[VB03] Vom Brocke, J. (2003): Referenzmodellierung - Gestaltung und Verteilung von Konstruktionsbegriffen, Berlin: Logos.

[Wa13] Walser, K. (2013): Development of an administration domain model based on a business architecture model for public administration. In: Sabucedo, L.A.; Rifon L.A. (Eds.): MeTTeG 2013 Proceedings of the 7th International Conference on Methodologies, Technologies and Tools enabling e-Government, University of Vigo, Spain: s.n.

[WR12] Walser, K.; Riedl, R. (2012): Zum Stand des kommunalen E-Government in der Schweiz. In: Eixelsberger, W. (Hrsg.): Kommunales E-Government. Linde Verlag, Wien, S. 97-118.

[WRoJ] Walser, K.; Riedl, R. (o.J): Skizzierung von Front- und Back-Office Architekturprinzipien des E-Governments, Bern: PEG – Kompetenzzentrum Public Management und E-Government, Berner Fachhochschule.

[ZBT01] Zaugg, R. J.; Blum, A.; Thom, N. (2001): Nachhaltiges Personalmanagement – Spitzengruppenbefragung in europäischen Unternehmungen und Institutionen. Bern: IOP-Verlag.

Anhang: Übersicht über das Gemeindeprozess-Referenzmodell

Aus Platzgründen wurde die Darstellung in einem Dropbox-Ordner abgelegt und kann dort entsprechend eingesehen werden. Die URL dazu lautet wie folgt:

https://www.dropbox.com/s/7p03y7aq0es13jr/Anhang%20Gemeindereferenzprozessmodell%20-%20Kopie.pdf?dl=0

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Digital Enterprise Architectures

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Digital Enterprise Architectures

Gunther Piller[1], Alfred Zimmermann[2], Kurt Sandkuhl[3], Rainer Schmidt[4], Ulrike Steffens[5] und Oliver Bossert[6]

Abstract: Neue Modelle für digitale Unternehmensarchitekturen mit Big Data, Services & Cloud Computing, mobilen Systemen, Internet of Things sowie Industrie 4.0 Ökosystemen machen eine enge Kooperation verschiedener Partner aus Wissenschaft, Anwendungsunternehmen, öffentlichen Organisationen, Softwarehersteller und IT- Dienstleister notwendig. Ziel dieser Zusammenarbeit ist die Zusammenführung neuer Konzepte und Möglichkeiten der Informationstechnologie zur bestmöglichen Unterstützung sich verändernder Unternehmensziele und -strategien. Software- und Unternehmensarchitekturen spielen hierbei eine zentrale Rolle. So werden Anforderungen bezüglich Flexibilität und Agilität in digitalen Unternehmen wesentlich durch serviceorientierte Ansätze unterstützt. Der Ordnungsgrad und die kosteneffiziente Gestaltung komplexer IT-Landschaften soll durch Digital Enterprise Architecture Management deutlich verbessert werden – passend zu neuen Möglichkeiten von Services & Cloud Computing, Big Data, sowie kollaborativen Geschäftsprozessen. Die Veranstaltung knüpft an die erfolgreichen Vorgängerworkshops auf den Jahrestagungen 2012-2014 der Gesellschaft für Informatik an. Zudem baut sie auf aktuellen Forschungsarbeiten aus der GI-Fachgruppe Software-Architektur und dem GI-Arbeitskreises Unternehmensarchitektur auf.

Programmkomitee

Karlheinz Blank – T-Systems
Oliver Bossert – McKinsey Frankfurt
Carsten Breithaupt – Deutsche Lufthansa
Michael Herrmann – Mercedes Benz Bank
Robert Hirschfeld – Hasso Plattner Institut
Ulrich Kalex – Software AG
Stefan Kühne – Universität Leipzig
Martin Mähler – IBM
Gunther Piller – Hochschule Mainz
Michael Pretz – Daimler
Karl Prott – Capgemini
Wilfried Reimann – Daimler
René Reiners – FIT
Ralf Reussner – KIT / FZI
Kurt Sandkuhl – Universität Rostock
Rainer Schmidt – Hochschule München
Christian Schweda – Hochschule Reutlingen
Ulrike Steffens – HAW Hamburg
Gottfried Vossen – Universität Münster
Alfred Zimmermann – Hochschule Reutlingen

[1] Hochschule Mainz, Lucy-Hillebrand-Straße 2, 55128 Mainz, gunther.piller@hs-mainz.de
[2] Hochschule Reutlingen, Alteburgstraße 150, 72762 Reutlingen, alfred.zimmermann@reutlingen-university.de
[3] Universität Rostock, Albert Einstein Str. 22, 18057 Rostock, kurt.sandkuhl@uni-rostock.de
[4] Hochschule München, Lothstraße 34, 80335 München, rainer.schmidth@hm.edu
[5] HAW Hamburg, Berliner Tor 5, 20099 Hamburg, ulrike.steffens@haw-hamburg.de
[6] McKinsey & Company, Taunustor 2, 60311 Frankfurt, oliver_bossert@mckinsey.com

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Impact of Cyber-Physical System Implementation on Enterprise Architectures: A Case Study

Kurt Sandkuhl[1]

Abstract: Cyber-Physical Systems (CPS) tightly integrate physical and IT (cyber) worlds based on interactions between these worlds in real time. The focus of this paper is on the question how CPS can be integrated the enterprise architecture (EA) and how CPS potentially affect the EA. The approach used in this paper is to evaluate a case study from transportation showing the development of a CPS. The contribution of the paper are (1) the description of a CPS case from an EA perspective, (2) an analysis of potential effects of the CPS on EA and (3) initial recommendations for CPS projects.

Keywords: Cyber-Physical Systems, Enterprise Architecture, Case Study

1 Introduction

Cyber-Physical Systems (CPS) tightly integrate physical and IT (cyber) worlds based on interactions between these worlds in real time [HG12]. CPS rely on communication, computation and control infrastructures commonly consisting of several levels for the two worlds with various resources as sensors, actuators, computational resources, services, humans, etc. The benefits and innovation potential attributed to Cyber-Physical Systems (CPS) is significant; as CPS are considered as the key to higher efficiency in many industrial domains [DBR14]. From an enterprise perspective CPS can contribute to product innovation (see [LS10] for an example from health industries), process innovation (see [Fis14] for an example from manufacturing) or business model innovation (see [SSS13]).

The focus of this paper is on the question how CPS can be integrated the enterprise architecture (EA) and how CPS potentially affect the EA. In general, an enterprise architecture captures the different conceptual layers of business model and IT in an enterprise, identifies the most important elements in these layers including their structural relationships, and shows the interdependencies between the layers. A frequently used approach in this area is TOGAF [OG14] which distinguishes between business architecture, application architecture, technology architecture and information architecture. Knowledge about potential effects of CPS on EA could be used during innovation projects for preparing CPS implementation.

The approach used in this paper is to evaluate a case study from transportation showing

[1] Universität Rostock, Institut für Informatik, Lehrstuhl für Wirtschafsinformatik, Einsteinstr. 22, 18059 Rostock, kurt.sandkuhl@uni-rostock.de

the development of a CPS. The contribution of the paper are (1) the description of a CPS case from an EA perspective, (2) an analysis of potential effects of the CPS on EA and (3) initial recommendations for CPS projects.

This work in progress paper is structured as follows: section 2 summarizes the background for our work. Section 3 briefly introduces the industrial case study. Section 4 shows an excerpt of the business processes and software architecture implemented for the CPS. Section 5 discusses the impact of the CPS on the existing EA of the case company. Section 6 summarizes the work and initial recommendations.

2 Background

This section summarizes the conceptual background for our work with focus on enterprise architecture management (2.1), Cyber-Physical Systems (2.2) and enterprise modelling (2.3).

2.1 Enterprise Architecture Management

Enterprise architecture (EA) denotes the fundamental elements and relationships of an enterprise in an appropriate model. Commonly, this includes the conceptual layers of business model and IT in an enterprise with their most important elements and structural relationships, including the interdependencies between the layers.. EA models have evolved over the last decade from pure IT architecture models into control instruments that can be used by the management as a tool for their business decisions and allow an integrated view on an enterprise. An EA supports the understanding and documenting of an organizational structure with all dependencies of artifacts and information objects necessary for business performance [BDMS10].

EA management (EAM) provides an approach for a systematic development of an enterprise's architecture in line with its goals by performing planning, transforming, and monitoring functions. The reasons for implementing an EA via EAM are manifold. On the one hand, it enables and supports the adaptation of IT to the business goals, the identification of problems or assistance coping challenges and on the other hand, it allows a detailed description of the conjunction between business and IT. This type of joint interaction results in creation of a common and consistent vocabulary for business processes and objects anon as well as for business functions and skills in the technical departments and the IT [Nie08].

2.2 Cyber-Physical Systems

As it was already mentioned, CPS tightly integrate heterogeneous resources of the physical world and IT world [Ant14]. This term is tightly related to such terms as Web

4.0 [ANF12] and Internet of Things [SM14]. Currently, there is a significant amount of research efforts in the area of cyberphysical networks and their applications, e.g., in production [Fis14], transportation [Wan14], and many other. Such systems rely on communication, computation and control infrastructures commonly consisting of several levels for the two worlds with various resources [TSLS14] as sensors, actuators, computational resources, services, etc. CPS belong to the class of variable systems with dynamic structures.

Having analysed the state-of-the-art of different CPS approaches and supporting technologies, among the other conclusions, Horvath and Gerritsen conclude that "the next-generation of CPSs will not emerge by aggregating many un-coordinated ideas and technologies in an incremental fashion. Instead, they will require a more organized and coordinated attack on the synergy problem, driven by an overarching view of what the future outcome should be" [HG12].

This means that the whole structure of the CPS to be implemented has to be built in advance based on the analysis of the required CPS functionality. The enterprise models can be a valuable information source in this case since they describe various aspects of an enterprise, acting units, their competences and relationships.

2.3 Enterprise Modeling

In general terms, enterprise modelling is addressing the systematic analysis and modelling of processes, organization structures, products structures, IT-systems or any other perspective relevant for the modelling purpose [Ver96]. Sandkuhl et al. [SSPW14] provide a detailed account of enterprise modelling approaches. Enterprise models can be applied for various purposes, such as visualization of current processes and structures in an enterprise, process improvement and optimization, introduction of new IT solutions or analysis purposes. Enterprise knowledge modelling combines and extends approaches and techniques from enterprise modelling. The knowledge needed for performing a certain task in an enterprise or for acting in a certain role has to include the context of the individual, which requires including all relevant perspectives in the same model [LK09]. A best practice for identifying these perspectives is the so-called "POPS*"-approach proposed by [Lil03]. POPS* is an abbreviation for the perspective of an enterprise to be included in an enterprise model: process (P), organization structure (O), product (P), systems & resources (S) and other aspects required for the modelling purpose (*). The best practice basically recommends to always include the four POPS perspectives in a model because they are mutually reflective: process are performed by the roles captured in the organization structure, the roles are using systems and resources which at the same time capture information about products; manufacturing and design of products is done in processes by roles using systems, etc. [LK09].

3 CPS Case Study

The CPS investigated in this paper was developed in an industrial project in transport and logistics industries. The logistics industry makes intensive use of modern information technology and CPS for achieving high efficiency of processes and solutions in a globalized market. One of the world's largest truck manufacturers designed and implemented new transport related services based on a CPS consisting of electronics in truck-trailers and an IT system using the information from the trailers for information logistics services. The trailers have a wireless sensor network (WSN) installed in the position lights. Each light carries a sensor node able to network with neighbouring nodes and furnished with a radar sensor. This sensor can be used for various purposes, including protection of the goods on the trailer against theft or surveillance of the trailer or its different compartments (e.g. by electronically sealing them). A gateway in the trailer is controlling the WSN in the position lights and communicates with the back-office of the owner of the trailer. Several services were developed within the project, which exploit the possibilities of combining sensor information and IT services. One of these services is additional protection of the trailer when parked against theft, also called secure trailer access control (STAC) (see [San12] for more details).

In this industrial case, the development of the STAC service and other information logistics services as part of the CPS followed a pragmatic approach including the following steps. These steps were started upon successful completion of technical feasibility studies regarding the technical infrastructure (i.e. sensor node – WSN – gateway – back-office communication):

- Business objectives: The enterprise management defined the business objectives to achieve. This included what actual services to offer in what priority, minimum number of customers, upper limit for investments into solution development and marketing, expected market share, and other general frame conditions.

- Business model: for each of the services, which were part of the business objectives, a business model was developed. For the business model of STAC, we used an approach from e-business proposed by Wirtz [Wir10] which separates the business model into partial models: capital model, procurement model, manufacturing model, market model, service offer model, and distribution model.

- CPS integration into the enterprise: as the business model does not describe how the new services will be integrated into the existing processes and structures of the enterprise delivering theses services, the next step was to design and integrate such structures and processes. This step was performed using enterprise modelling techniques (see section 2.2) and will discussed in detail in section 4.

- Specification of CPS: the specification of the CPS has to include the architecture and functional/non-functional characteristics of the technical solution and the set of operational service for managing and operating the technical solution. In section 4.3 we will focus on a selected aspect of the technical solution only: the architecture of the back-office solution

4 CPS Business Processes and Software Architecture

As discussed in section 3, an enterprise model was developed for describing the integration of the CPS into the enterprise and a software architecture as a basis for designing the actual technical system. As the enterprise model forms the basis for changes in the business architecture and the software architecture is relevant for the application, both will be briefly described in this section.

4.1 CPS integration into the enterprise

The business model developed for the CPS indicates where integration of the new service into the existing structure of the enterprise is needed. The manufacturing model defines what parts of the service are implemented in-house, i.e. from here we identify the existing roles and organization units (e.g. i.e. service operator, infrastructure operator, project manager) and resources involved (e.g. (contract management, configuration environment). Procurement model and service offer model help to identify external roles, external processes and product structure elements. However, the business model is neither precise enough for specifying the exact implementation in the enterprise nor meant for this purpose. Thus, we extended the enterprise model of our case company for the new service, which also visualizes how the CPS is integrated into business processes.

Fig. 1 shows an excerpt from the enterprise model for STAC. In the middle of the figure, the high level business process “STAC service” is depicted, which consists of “configure service”, “perform service” and “discontinue & report”. For all activities, several refinement levels exist. The figure only shows the first refinement level for the second activity “perform service”. On the lower left of the process container, the roles involved are modelled (including internal and external roles); on the lower right of the process container, the systems and infrastructure resources are shown. Above the process container, important documents (e.g. contractual elements, like terms and conditions) are modelled. The arrows in the figure are typed relations between the model elements, which for example show which role “is_responsible” for what activity in the process. Both, human resources and infrastructure resources are part of specific sub-models for the overall enterprise not shown in the figure.

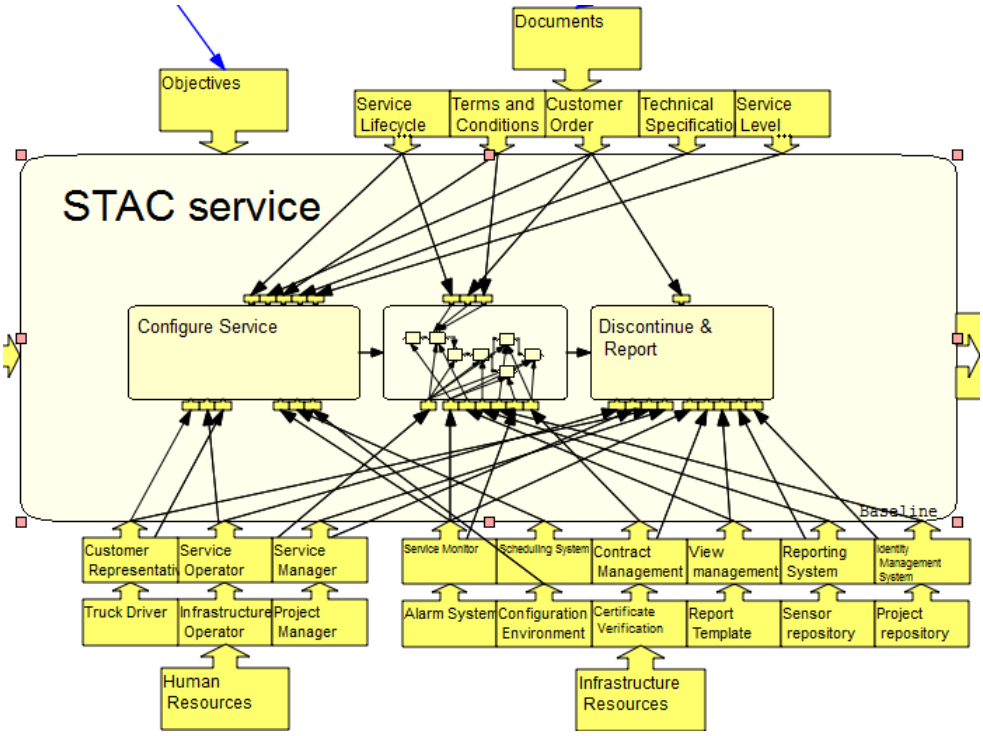


Fig. 1. Enterprise model excerpt for STAC with focus on process/role/resource

4.2 Software Architecture

Figure 2 depicts the components of the general software architecture defined for the STAC service. Grey-shaded rectangles denote components newly developed for STAC, rounded rectangles represent existing IT-systems from customers or suppliers to be integrated, white shaded rectangles depict existing enterprise-internal systems to be integrated. Rectangles with dotted lines indicate that the component is optional and will be implemented in a later version of the system. The arrows indicate the direction of information flow between these systems.

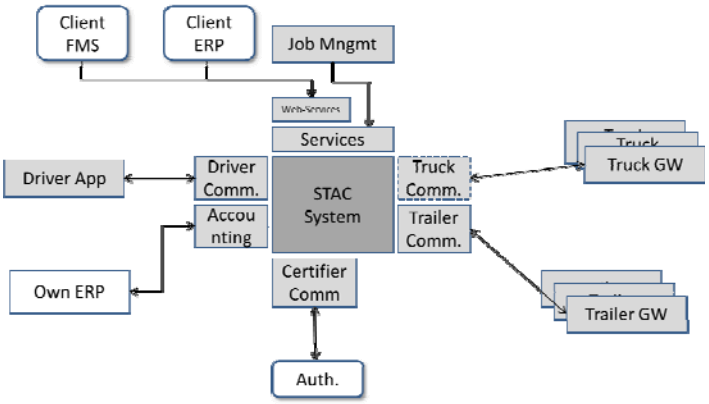


Fig. 2: STAC software architecture

The core of the software architecture is the STAC application, which manages all trailer configurations and connections, the ongoing trailer surveillances, the authentication of drivers and the management of alarms. The STAC application has interfaces to the enterprise internal ERP system for registering the performed accounting-relevant activities, and the communication with driver, trailers and (optionally) trucks. The interfaces to external applications concern the fleet management systems (FMS) of clients in order to receive basis information about the trailers to be supervised and their assignments, the clients' ERP system for receiving information about drivers and accounting and the trust center for driver authentication. Furthermore, there is a specific job management application for registered clients which allows for status check of ongoing surveillance assignments and of changes in / new creation of assignments.

5 CPS effects on EA

When investigating how the CPS described in section 4 can be integrated in the enterprise architecture of the transportation company and how the CPS in turn affects the EA we use the TOGAF layers, which include business architecture, application architecture, information architecture and technology architecture, i.e. we discuss the integration and effects layer by layer.

In the business architecture layer, the CPS is represented in the first place with the new capabilities the enterprise has due to the CPS. These capabilities are expressed with new functions implemented in the enterprise, business services offered to the customers and business processes implemented for providing these business services. Some of the new services were already discussed in section 3 (e.g. electronic fence or STAC), an excerpt of the business processes are shown in the enterprise model in section 4.1. Besides adding new "value creation" elements to the business architecture layer, the business services implemented by the CPS also need supporting services which already are

existing, like customer care, accounting or marketing. Since the CPS offers services which can be considered a completely new business model, even the target group is different from existing business lines, which requires new sections of customer care and marketing within the established organizational structures. However, the existing accounting capabilities can also be used for the new CPS services.

From the perspective of the application architecture, a number of new applications were added which are specific for the new services, like the core of the STAC application which controls the data traffic to the gateways in the trailers or implements the control flow for driver authentication or alarming security service providers in case of assumed trailer theft. But there are also some applications which potentially could be shared with other business areas of the enterprise. One example is the exchange of accounting information with the FMS systems of the clients or with the clients' ERP systems. This exchange is done using a message-based middleware, like IBM's MQ series, which can also be used for integrating some of the internal accounting systems. For the CPS integration it was decided to initially use a separate message queuing middleware installation as the volume of message to be transferred was difficult to estimate and as this middleware potentially also could be used as part of the infrastructure in trailer communication. Regarding the information architecture, the initial idea was to use the established sales and delivery, customer care and accounting information structures for the new CPS based services, which also would mean that the applications for managing this information could be used. However, it quickly showed that the parameters characterizing a "product" using the CPS were quite different from the established products in the enterprise. Thus, a new instance of the sales and delivery application had to be installed configured for the specific business related to the CPS whereas the customer care and accounting could be used with only additional configurations.

On technology architecture level, the situation was similar to the application architecture: on the one side there a new components and even a completely new infrastructure which has to be implemented (i.e. trailer gateways, communication links to the back-office, scalable compute servers for processing of events from the trailers, etc.). On the other side, some of the back-office functionality is suitable for deploying them on existing resources in the computing center of the enterprise. In this CPS case, it was decided to deploy the additional middleware services mentioned in the application architecture on already existing resources, as the capacity seemed to be sufficient.

6 Summary and Conclusions

Based on a CPS example from transportation industries, the paper discussed changes in the EA implied by the CPS. Although this is case study is just one example for CPS systems, three recommendations can be derived from the experiences collected. These recommendations have to be subject to validation in future projects:

Recommendation 1: design the first version of the CPS architecture without taking into account the existing EA and build a prototype to gather experiences regarding critical issues of the CPS. In a second version identify overlaps between existing EA and CPS architecture and analyze potential synergies (e.g. reusable applications or scalable technology components).

Recommendation 2: During analysis of EA, pay specific attention to functional identical areas, performance, security and reliability requirements.

Recommendation 3: In the technology architecture, separate the infrastructure for control and operations of the CPS from the infrastructure for other enterprise applications.

References

[ANF12] Aghaei, S.; Nematbakhsh, M.; Farsani, H.: Evolution of the World Wide Web: From WEB 1.0 TO WEB 4.0. International Journal of Web & Semantic Technology, 3(1), 1-10, 2012.

[Ant14] Antsaklis, P.: Goals and Challenges in Cyber-Physical Systems Research Editorial of the Editor in Chief. IEEE Transactions on Automatic Control, 59(12), 3117-3119, 2014.

[BDMS10] Buckl, S.; Dierl, T.; Matthes, F.; Schweda, C.: Building Blocks for Enterprise Architecture Management Solutions. In: Practice-Driven Research on Enterprise Transformation, LNBIP Volume 69, pp 17-46, Springer, 2010.

[DBR14] Deutsche Bank Research: Industry 4.0 - Upgrading of Germany's industrial capabilities on the horizon. Available at http://www.dbresearch.de/PROD/DBR_INTERNET_EN-PROD/PROD0000000000333571/Industry+4_0%3A+Upgrading+of+Germany%E2%80%99s+industrial+capabilities+on+the+horizon.pdf, 2014.

[Fis14] Fisher, A.; Jacobson, C.; Lee, E.; Murray, R.; Sangiovanni-Vincentelli, A.; Scholte, E.: Industrial Cyber-Physical Systems - iCyPhy. Complex Systems Design & Management, pp. 21-37. Springer International Publishing, 2014.

[HG12] Horvath, I.; Gerritsen, B.: Cyber-Physical Systems: Concepts, Technologies and Implementation Principles. In Horvath, I., Rusak, Z., Albers, A. Behrendt, M. (eds.) Proceedings of TMCE 2012, 19-36, 2012.

[Lil03] Lillehagen, F.: The Foundations of AKM Technology, 10th International Conference on Concurrent Engineering (CE) Conference, Madeira, Portugal, 2003.

[LK09] Lillehagen, F.; Krogstie, J.: Active Knowledge Modelling of Enterprises. Springer, ISBN: 978-3-540-79415-8, 2009.

[LS10] Lee I.; Sokolsky O.: Medical Cyber Physical Systems. Proc. IEEE Design Automation Conference, Proc 47th Design Automation Conference , Anaheim, CA, USA, pp. 743-748, 2010.

[Nie08] Niemi, E.: Enterprise architecture benefits: Perceptions from literature and practice. Evaluation of enterprise and software architectures: critical issues, metrics and practices. University of Jyväskylä, IT Research Institute, ISBN 978-951-39-3108-7, 2008.

[OG14] The Open Group: TOGAF® Version 9.1, http://pubs.opengroup.org/architecture/togaf9-doc/arch/chap32.html, (accessed 08.05.2015), 2014.

[San12] Sandkuhl, K.; Borchardt, U.; Lantow, B.; Stamer, D.; Wißotzki, M.: Towards Adaptive Business Models for Intelligent Information Logistics in Transportation. In: BIR 2012, Nizhny Novgorod, Russia, September 24-26, 2012. ISBN 978-5-502-00042-0, 2012.

[SM14] Skarmeta, A.; Moreno, M.: Internet of Things. Secure Data Management, pp. 48-53. Springer International Publishing, 2014.

[SSPW14] Sandkuhl, K.; Stirna, J.; Persson, A.; Wißotzki, M.: Enterprise Modeling: Tackling Business Challenges with the 4EM Method. Springer. ISBN 978-3662437247, 2014.

[SSS13] Smirnov, A.; Sandkuhl, K.; Shilov, N.: Multilevel Self-Organisation and Context-Based Knowledge Fusion for Business Model Adaptability in Cyber-Physical Systems. Manufacturing Modelling, Management, and Control. Saint Petersburg, Russia, Jun. 19, 2013. Volume # 7: Elsevier, IFAC (IFAC proceedings volumes), pp. 609-613, 2013.

[TSLS14] Teslya, N.; Smirnov, A.; Levashova, T.; Shilov, N.: Ontology for Resource Self-Organisation in Cyber-Physical-Social Systems. 5th International Conference of Knowledge Engineering and the Semantic Web (KESW 2014), CCIS, 468, pp. 184-195. Springer, 2014.

[Ver96] Vernadat, F. B.: Enterprise Modelling and Integration. Chapman & Hall, 1996.

[Wan14] Wan, J.; Zhang, D.; Zhao, S.; Yang, L.; Lloret, J.: Context-Aware Vehicular Cyber-Physical Systems with Cloud Support: Architecture, Challenges, and Solutions. Communications Magazine, IEEE, 52(8), 106-113, 2014.

[Wir10] Wirtz, B.: Electronic Business. Gabler Verlag, 2010.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

A Decision-Making Case for Collaborative Enterprise Architecture Engineering

Dierk Jugel[1,2] , Stefan Kehrer[1], Christian M. Schweda[1] and Alfred Zimmermann[1]

Abstract: In modern times markets are very dynamic. This situation requires agile enterprises to have the ability to react fast on market influences. Thereby an enterprise' IT is especially affected, because new or changed business models have to be realized. However, enterprise architectures (EA) are complex structures consisting of many artifacts and relationships between them. Thus analyzing an EA becomes to a complex task for stakeholders. In addition, many stakeholders are involved in decision-making processes, because Enterprise Architecture Management (EAM) targets providing a holistic view of the enterprise. In this article we use concepts of Adaptive Case Management (ACM) to design a decision-making case consisting of a combination of different analysis techniques to support stakeholders in decision-making. We exemplify the case with a scenario of a fictive enterprise.

Keywords: Enterprise Architecture, Adaptive Case Management, Analysis

1 Motivation

In modern times markets are very dynamic. Enterprises have to be agile to be able to act fast on changing influences. In particular, their business models are highly volatile and have to be adapted. Changing business models may have several impacts on the enterprise architecture (EA) including business processes, business units, information systems, and IT infrastructure. These architectural elements have manifold relations to each other and form an EA to a highly complex structure. Enterprise Architecture Management (EAM) is a method to support stakeholders in the adaptation and transformation processes. Based on an up-to-date description of the current EA, analyses are performed to understand adaptation needs and implications.

In practice, stakeholders commonly use visualizations to analyze the EA. These visualizations, i.e. views on the EA, is usually created using EAM tools. However, as described in [Ma08] and [RZM14], these tool-created views are often report-like, i.e. static with respect to the displayed information. A more powerful support for EA analysis can be built on the paradigm of the visual analytics, as outlined by Keim et al.

[1] Reutlingen University, Herman Hollerith Zentrum, Reutlingen, Germany, {dierk.jugel, christian.schweda, alfred.zimmermann}@reutlingen-university.de, stefan.kehrer@student.reutlingen-university.de
[2] Rostock University, Rostock, Germany, dierk.jugel@uni-rostock.de

[Ke08]. For visual analytics, algorithmic analyses performed by the tool are semi-automatically and interactively applied to the structure under consideration. For the field of EAM, this method can be regarded a novelty, as a recent survey on the state-of-the-art of visual EAM conducted by Roth et al. [RZM14] does not identify any related concept in today's EAM tools.

Buckl et al. [BMS09] classify EA analysis techniques described in literature. They identify three types of analysis techniques: (1) expert-based, (2) rule-based, and (3) indicator-based. They further indicate that expert-based analyses are central to the design processes in EAM and are complemented by techniques of the two other kinds, providing hints and indications. Buckl et al. discuss the downsides of expert-based analyses [BMS09] being time-consuming, error-prone, and dependent on the availability of expert knowledge at the organization's disposal.

Rule-based and indicator-based techniques are more formal in their nature and can be described as algorithms. Therefore, these techniques are automatable, repeatable and independent of individual experts. As described in [BMS09] such analyses can be implemented via pattern matching (rule-based) and via aggregation operations (indicator-based).

In this paper, we explore how expert-based, rule-based, and indicator-based analysis techniques can be integrated into a comprehensive process for EA analysis and explore how this process contributes to a collaborative decision-making process. We build this process, among others, on the groundwork provided by Jugel et al. [JSZ15] for collaboratively performing visual analytics. We integrate this groundwork with existing work on EA analysis, revisited in Section 2.1. Further, we explore techniques for collaborative decision-making (cf. Section 2.2) and reflect on the analysis and decision-making processes being knowledge-intensive processes (cf. Section 2.3). In Section 3, we present a process for EA analysis integrated into a process for collaborative decision-making in EAM. The process' underlying meta-model builds on the meta-model provided by Jugel et al. [JSZ15] and its mechanism for annotating EA models with additional information. We exemplify the approach in Section 4 by describing an application case. Section 5 concludes the article and gives an outlook in topics of future work.

2 Related Work

In this section, we revisit related work in the field of EA analysis techniques, collaborative decision-making, and adaptive case management. In Section 2.1 we start with presenting different analysis techniques. Afterwards we address collaborative decision-making in Section 2.2 and finish in Section 2.3 with related work in adaptive case management.

2.1 EA Analysis Techniques

Buckl et al. [BMS09] identify three types of EA analysis techniques, on which we will subsequently reflect. We start with expert-based techniques described in literature.

Jugel et al. describe in [JS14] an interactive cockpit approach to support stakeholders in EA planning. The authors identify several interactive functions that are useful in this field. Two interactive functions named “graphical highlighting” and “graphical filtering” are of interest. These functions enable stakeholder to annotate architecture elements that are of interest or not. The annotation mechanism described provides adding notes and grouping elements. In addition, visual variables can be assigned to highlight architecture elements, which are of interest or to filter elements, which are not of interest. These functions are part of expert-based analysis techniques, because stakeholders are doing the analysis by hand without formalization.

Whereas expert-based techniques are dependent on expert knowledge, rule-based techniques can be described as algorithms to support automated analysis execution [BMS09]. Hanschke provides a rule-based approach to EA analysis with the so-called analysis patterns presented in the appendix A of [Ha13]. These analysis patterns are described as practice-proven and generalized templates to find needs for action and potential improvements concerning the EA. Hanschke identifies five different categories: 1) redundancy, 2) inconsistency, 3) organizational need for action, 4) implementing business requirements and 5) technical need for action and potential improvements. Each analysis pattern is structured using a canonical form including the following characteristics: id, name, version, description, context, dependencies, result, and example. Each pattern provides a textual prescription on how to identify shortcomings and derive improvements in the architecture. This prescription gives guidance to an experienced enterprise architect, but is not translated to a formal, i.e. algorithmic, manner. Whereas these analysis patterns cannot be automated directly, Jugel et al. address this issue in [Ju15] and propose an automated analysis technique to increase the level of formalization. Thereby the authors use a generic annotating mechanism to enrich EA models with additional knowledge emerged in discussions and performed analyses during decision-making processes [Ju15].

Ramos et al. propose in [Ra14] a characterization of analysis functions including a specification of the algorithms. Therefore, the authors perform a rule-based analysis technique. The structure for each analysis function contains the following information: name, description, dimension, type, layer, entities and relations, structural attributes and algorithm. An algorithm describes the information extraction of the analysis function from the model. The analysis process itself is depicted as a cyclic procedure of querying and enriching the EA model until the result is achieved. The authors are currently working on a catalog of so called analysis functions. We consider their approach to be similar to the concept of Hanschke’s analysis patterns with a stronger focus on implementation.

Indicator-based analysis techniques can be implemented via aggregation operations. Matthes et al. [Ma11] present quantitative, metrics-driven EA analyses by quantitatively assessing architectural properties and therefore use an indicator-based analysis technique. The proposed KPIs provide necessary measurement capabilities for EAM, needed to aid planning and controlling the EA. The EAM KPI Catalog contributes in presenting ten common EA management goals and 52 KPIs including the underlying EA model. Each KPI is described by the following characteristics: description, underlying EA model, goals, calculation, code, sources, organization-specific instantiation and the affected layers of the EA. The EAM KPI Catalog also provides a good basis for implementation purposes. Especially the calculation section describes the algorithm for calculation of each KPI concisely. While KPIs can be used to detect need-for-action, optimization potentials in the EA itself are not identified. As stated by Buckl et al. the results of indicator-based analysis techniques have to be interpreted carefully [BMS09]. An experienced enterprise architect would be needed to identify the elements of the EA causing the value of a KPI. However, the catalog only consists of KPIs that are defined in COBIT [IT07]. KPIs like the number of applications in operation or used technologies are not part of this catalog. Legner et al. mention such measurements [LL12]. The authors state the importance of suitable KPIs to monitor the EA and related activities. To implement KPIs they introduce a so-called "EAM cockpit" that is a multi-dimensional KPI system that covers three important dimensions: impacts in business terms, status of the current EA and EAM adoption in the organization. Each dimension includes several KPIs. For instance, the dimension "EA status" includes the KPI "Total instances" of particular types within an EA. In addition, the authors relate the dimensions with stakeholders, which are interested in. For instance, enterprise architects are interested in KPIs of the dimension "EA status". The presented KPI system is very useful to introduce an EA monitoring. However, in contrast to [Ma11] the authors do not describe calculations and the KPI's underlying metamodel to perform them.

2.2 Collaborative decision making

Making good decisions concerning the design and evolution of the EA is key to the field of EAM. Johnson et al. recommend in [JE07] a goal-driven approach for EAM. The authors especially address how to derive decision-relevant information from EA models. Johnson et al. discuss that architecture-related goals have to be operationalized to provide a foundation for the decision-making processes. Therefore, the authors propose a set of activities related to decision-making: (1) The decision maker must settle on a goal or success criterion, (2) Alternative designs have to be identified, (3) Effects of the alternative designs on the goals must be elicited, (4) The decision-maker needs to decide on what detail information to collect with respect to the different design alternatives, (5) Information needs to be collected, (6) Collected information has to be consolidated into an aggregated assessment, (7) Finally, the decision is taken.

These activities underline the need of a clear decision-making process to make good decisions concerning architectural efforts. Johnson et al. do not focus on how the EAM organization is structured [JE07]. Whereas these outlined activities can be aggravated by disruptive factors such as unclear goal definitions, a lack of expert knowledge or uncertain information in a hierarchical organization, these activities can be much more difficult in a collaborative EA environment. The authors stress the specific creation of viewpoints to support stakeholders in their decision-making tasks [JE07].

EAM uses models and visualizations of relevant information to support stakeholders in their collaborative tasks [Ma08]. The models are used in describing, documenting and sharing relevant knowledge needed for decisions in a collaborative EA environment. Many EA modeling approaches advocate collaborative modeling concepts and their advantages. Sandkuhl et al. propose a workshop approach for participatory modeling [Sa14]. The authors deem the involvement of stakeholders with the best knowledge needed to reach a particular workshop goal as valuable. Additionally, stakeholders' involvement increases the acceptance of created models [Sa14].

Hamm et al. combine the aforementioned approach of Johnson et al. [JE07] with collaborative modeling techniques to reach a model-based, goal-oriented decision-support for EA decision-making [HK15]. The authors utilize an enhanced version of the ArchiMate Motivation extension to cover goal-oriented information demands in collaborative EA environments [TOG12]. The presented approach focuses on strategic-alignment in collaborative EA decision-making processes and omits procedural aspects as EA analysis, which establishes an information base for EA decision-making [HK15].

Plataniotis et al. describe an approach focusing on ex-post modeling of EA decisions [PDP14]. The presented metamodel contains new concepts not covered by ArchiMate [TOG12]. The approach of Plataniotis et al. has some limitations: (1) the authors assume that a single stakeholder takes the decisions. In practice, many stakeholders are involved in decision-making. (2) The approach's intention is an ex-post documentation of decisions. This is not adequate particular in a collaborative EA environment. (3) An EA Decision is associated with exactly one layer of ArchiMate. This hinders the documentation of complex interdependencies within two or more EA layers. Whereas Plataniotis et al. contribute in presenting a metamodel for ex-post modeling EA decisions, collaborative and process-oriented aspects are left [PDP14]. We consider procedural information in complex decision-making as an integral part, especially in collaborative EA environments. This enables stakeholders to retrace and understand decision-making processes and resulting decisions.

Jugel et al. [JSZ15] enhanced the approach of Plataniotis et al. [PDP14] with collaborative aspects. Their decisional metamodel for collaborative EA engineering enables real-time modeling of information concerning EA decisions, and thereby a retraceable documentation of decision-making work. The presented metamodel contains only a few elements to fuel practicability through reduced modeling overhead [JSZ15].

Whereas stakeholder can document information for EA-decision-making during an expert-based analysis process (e.g. related viewpoints), procedural aspects of EA decision-making are not covered by the developed metamodel.

The upcoming standard Decision Model and Notation (DMN) of OMG [Om14b] discern three usage models: for modeling human decision-making, for modeling requirements for automated decision-making, and for implementing automated decision-making. DMN bridges the gap between business decision designs and their implementation by providing a common notation for decision models. The purpose of DMN is to facilitate a decision model framework, which is easily usable for decision diagrams and as a base for optionally automating decisions. Decision-making support is addressed from basically two perspectives: normal BPMN business Process Models can be expanded by defining specific decision tasks, or decision logic can be used to support individual decisions, e.g. business rules, decision tables, or executable analytic models. DMN can additionally provide a third perspective to bridge between business process models and decision logic by introducing the Decision Requirements Diagram. Complementary to the DMN notation, which is used to model decisional relationships and concepts like Decision, Input Data, Business Logic, Application, Application Risk, etc. DMN introduces an expression language to represent decision tables, decision rules, and function invocations. Today we are exploring the suitable usage and close link of DNM for decisional support logic within our architectural engineering and analytics research.

2.3 Adaptive Case Management in EAM

Adaptive Case Management (ACM) [Sw10] and [Sw13] offers a lightweight model to support knowledge-intensive processes, which are driven by user decision-making. Knowledge processes of usually high-skilled stakeholders, like enterprise architects, require process adaptations at run-time. ACM is not dictating a predefined course of action [HPM14] and provides the necessary information and knowledge support to be able to solve a case. A case [Sw13] is typically a collection of all relevant information into one place, which is handled by one or more knowledge workers during solving this case. The case is the jointly used focal point for assessing the situation, initiating activities and processes, implementing the work, and reflecting results based on a history record about what was really done. A case brings together all the necessary resources and also tracks everything that has happened into a record history, which can be mined to synthesize best practices, patterns of success, and used and extended instruments.

Fundamental aspects and requirements, which are relevant for ACM, are mentioned in [HPM14]:

1. The adaptation aspect of ACM consists of content, people, and reporting capabilities to be able to change the knowledge process at run-time by end-users. Additionally to the adaptation aspect a knowledge worker should be able

to continuously improve his case templates.

2. The organization aspect groups policies, processes, and data. In ACM data is the dominant factor as opposed to the process-oriented view from BPM. Knowledge work requires the integration of data [HPM14] into the execution process.

3. The case handling aspect is about collaboration, decision support, and integration of resources, events, and communication. Complex problems are typically solved collaboratively by involving individual stakeholders in respect of different necessary knowledge types and stakeholder concerns. Decision support requires transparency within a shared understanding of analyzed EA scenarios by named stakeholders.

Opposed to routine work, which can be supported by business process management because of its repeatable kind, knowledge work is typically unpredictable. Knowledge workers [Fi11], [Fi14] are acting under uncertainty. An unpredictable process [Sw13] does not repeat in routine patterns and emerges as the work is done. The practice of preparing for many possible courses is called agility.

Differentiating seven domains of predictability [Sw13] case management can be focused on two main types:

1. Product Case Management: Supports design-time knowledge processes with a well-known set of actions, having much variation between individual cases. It is not possible to set out a single fixed process. Knowledge workers are actively involved in deciding the course of events for a case.

2. Adaptive Case Management: Knowledge workers are involved not only in the case, and picking predefined actions, but they are constantly adapting the process and striving for innovative approaches, and may want to share and discuss process plans.

The Object Management Group (OMG) has published the Case Management Model and Notation (CMMN) [Om14a] as a first step to support modeling for case management scenarios. In [Ha14] was implemented a case study of a TOGAF-style process for EAM with CMMN.

3 Collaborative decision making

In this section we adapt our approach presented in [JSZ15] by supporting collaborative decision-making processes in more detail. Firstly, we present a collaborative decision-making case. Afterwards we adapt the collaborative decision-making metamodel of [JSZ15]. Decision-making processes are knowledge-intensive processes that strongly

depend on stakeholders, which participate on. Thus specifying a rigid structure, like the classical business process management does, for decision-making processes is impossible. For that purpose, we regard adaptive case management (ACM) to be more suitable. ACM focuses on cases instead of processes, which reflect decision-making activities. The Case Management Modeling Notation (CMMN) [Om14a] provides flexible processes, like "DiscretiaryItems" to reflect optional tasks. We use this notation to model a collaborative decision-making case to refine EAs. Figure 1 illustrates this case.

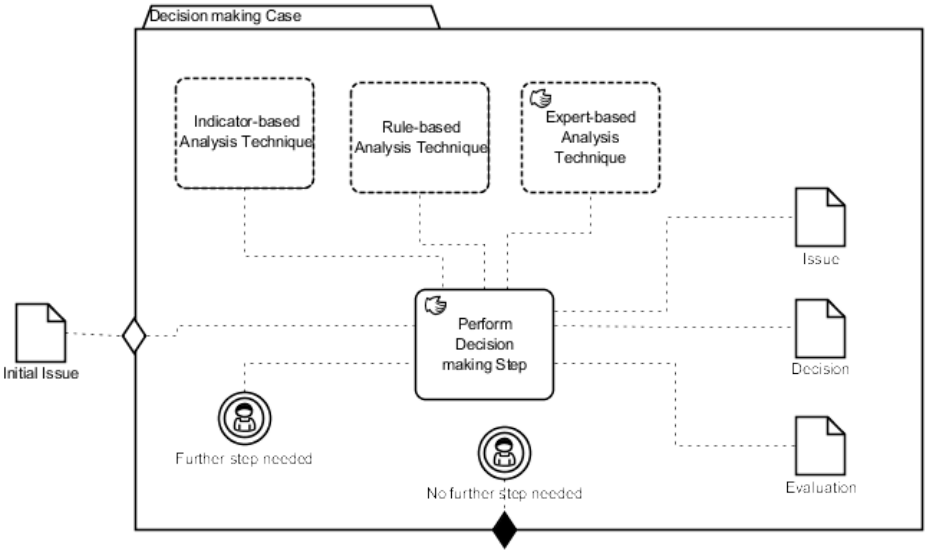


Fig. 1 CMMN model of collaborative decision making case

We regard an “issue” to be the starting point of a collaborative decision-making case. This issue describes the problem space of the decision-making activity, which aligns with the perspective of Mayring [Ma10]. We further assume that goals and success criterions, as required by Johnson et al. [JE07], have already been defined as part of strategic management activities. The issue is the reason why an EA has to be analyzed and a refinement thereof is probably needed. Based on this issue, involved stakeholders have to choose viewpoints they need to analyze the issue and quantitative as well as qualitative analysis techniques are applied to generate additional insights [Ma10]. The stakeholders perform a decision-making step that accounts for and interprets these additional insights [Ma10].

In the decision-making case, described in Figure 1, the central activity is the *decision-making step.* In this step, the stakeholders can employ different kinds of available analysis techniques – indicator-based, rule-based and expert-based – to obtain additional

insights. Analysis techniques that can be chosen are predefined and part of a catalog. The catalog is independent of a particular case. Each decision-making step is based on case data consisting of an EA model and additional insights elicited in previous steps. Consequently, the insights gained during each step contribute to the "case file" of the decision-making case. Values of KPIs, for example, that derive from existing information in the EA model are used to graphically highlight architecture elements performing poor with respect to the analyzed quality. The stakeholders use these highlights in their interpretation that concludes a particular decision-making step. The interpretation yields further items in the case data:

- An evaluation represents the stakeholder's opinion on the analysis results.
- A new issue refines the previously analyzed one based on the analysis.
- A decision reflects an actual design that is useful to resolve the issue.

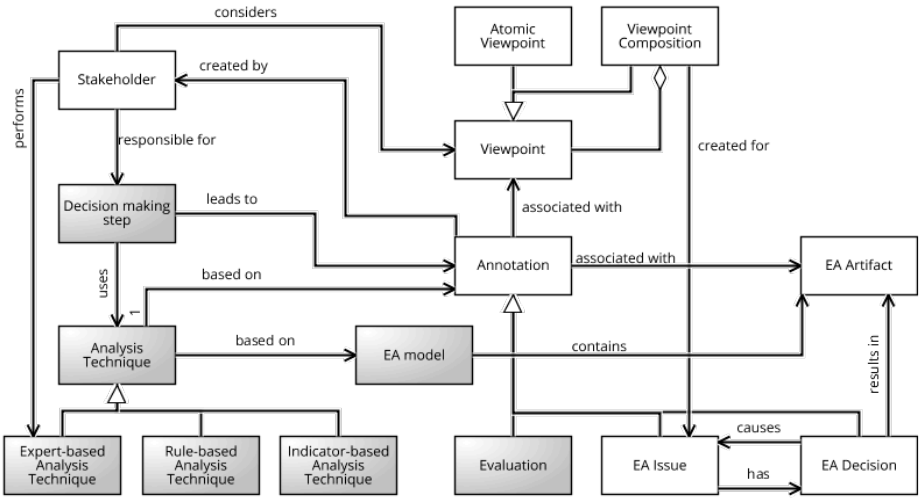


Fig. 2 Collaborative EA Decision Making Metamodel

An activity described by Johnson et al. [JE07] is the identification of alternatives, which could be reflected in different decisions. In the final step of the decision-making process, not all previously evaluated designs will prevail. At the end of every decision-making step, the stakeholders have to choose, whether additional information is required or not – represented by to "UserEventListeners" in the CMMN diagram in Figure 1. The case file of the decision-making case has to be structured appropriately to accommodate for the described decision-making process. We propose a structure for the case file that builds on the collaborative decision metamodel presented in [JSZ15]. The metamodel focuses on the documentation of decision and rationalizing information. The extended metamodel from Figure 2 incorporates also the decision-making techniques to retain all

information needed to review a decision case after the fact. In this sense, the concepts presented in the metamodel can be regarded as definitions for case file items in the terminology of CMMN. Figure 2 illustrates the refined collaborative decision- making metamodel. Changed or added concepts of the metamodel described in [JSZ15] are visually highlighted.

The *Annotation* is the key concept of the metamodel and represents additional information gained during the decision-making case. In addition, *Annotations* are the triggers for the next *Decision Making Step*. One or more *Stakeholders* are responsible for a step and perform them. Within *Decision Making Step* stakeholders can choose between different analysis techniques to get additional information needed to satisfy their information demands. *Analysis Techniques* are based on *Annotations* as well on the *EA model*. The *EA model* consists of *EA Artifacts* of an EA, *Annotations* describe additional information related to *EA Artifacts*. In contrast to the metamodel described in [JSZ15], we adapt the refinements of *Annotation*. The metamodel of [JSZ15] distinguishes between *Detailed Information* and *Task*. While *Detailed Information* is defined as any additional information that is relevant for decision-making, a *Task* describes seeking needed information, which is not part of the *EA model*. We redefine the concept of *Detailed Information* by the concept of *Evaluation*. Thus, a *Decision-making step* encapsulates an *Analysis Technique* and the interpretation thereof by responsible stakeholders the gained information of a step conforms to an *Evaluation*. Consistent with the process from Figure 1, we omit the *Task* concept and interpret the task as an *Expert-based Analysis*. Such analysis describes all activities to analyze the model without resorting to predefined algorithms and rules. In particular, activities for collecting information that is not contained in the EA model are expert-based analyses. *Stakeholders* consider *Viewpoints* to satisfy their information demands. These viewpoints represent information from the EA model as well as additional information present in the form of annotations. The concepts of *EA Issue* and *EA Decision* are used in line with Plataniotis et al. [PDP14], but we discuss only the high-level relationships between these concepts.

4 Collaborative analysis & decision making scenario

In this section, we present a scenario of a fictional enterprise to exemplify the collaborative decision making case described in Section 3. We demonstrate the combination of different analysis techniques we described in Section 2.1 that form a decision making case.

The CIO of a large enterprise tells the enterprise architect John that the IT budget has to be reduced. In the first decision-making step John has the task to consider how cost saving potentials can be detected. Based on his experience John knows that information systems and technical components occasion much costs. Thus, John does not use any

formalization it is an expert-based analysis technique. Based on his finding that information systems and technical components occasion much costs, he involves Mark, the application responsible, and Joe, a technology expert. They meet at a cockpit, like [JS14] describes. The cockpit consists of several screens, which can be used to display coherent viewpoints simultaneously.

First, they choose viewpoints that fulfill their information demands to analyze the issue. They use different types of visualizations described by [Ha13], p. 87. They choose the following viewpoints:

- Information System Overview Viewpoint: This viewpoint gives an overview about information systems. For reasons of complexity, they decide only displaying the names of each information systems.
- Technical Component Overview Viewpoint: In analogous of the Information System Overview Viewpoint this viewpoint includes the name of each technical component
- Business Support Map: This viewpoint provides information about which information is used by which business for which business process.
- Technical Development Cluster: This viewpoint provides information about which technical component is used by which information system.

They start analyzing the information systems by using the “Information Systems Overview Viewpoint” and the KPI "Total instances of information systems" (c.f. [LL12], p. 187) to get an overview about the situation. A KPI corresponds to an indicator-based analysis technique defined by [BMS09]. The value of the KPI is at a normal level, so they decide to do nothing for now by documenting a corresponding decision. Afterwards the next *Decision making step* focuses on analyzing technical components. For this purpose, they focus on analyzing the “Technical Component Overview Viewpoint” by using the indicator "Total instances of technical components" (c.f. [LL12], p. 187) to get information about the total number of technical components. Joe evaluates the number as very high. Therefore, further information about technical components is needed. They create an issue named “Too many technical components”. Joe suggests using the rule-based analysis technique "R-T-TB Redundancies in technical development" (c.f. appendix A of [Ha13], p. 76) to get redundant technical components, which are candidates to retire. The analysis technique identifies the following technical components as potentially redundant, because all of them are database management systems: TC1, TC2, TC3, TC4, TC5 and TC6. These candidates are highlighted in the “Technical Component Overview Viewpoint” and “Technical Development Cluster” with a red fill-color. Now the stakeholders can easily identify the candidates in different contexts of the viewpoints. They create an evaluation consisting of these six technical components to document the redundancy of database management systems.

To limit the number of candidates John suggests using the rule-based analysis technique "T-IS/T/B-TZ Technical State of Technical Components" (c.f. appendix A of [Ha13], p. 67) to get information about their state of health. In his opinion it is of advantage if they retire technical components, which have a bad state of health. The result of the analysis technique is shown by different fill-colors depending on their calculated state of health. To avoid an information overflow all technical components, which are not identified as redundant are hidden in the “Technical Component Overview Viewpoint”. Result is that TC1, TC2 and TC3 have a bad state of health. They document this interpretation by creating a corresponding evaluation. Mark notes that they have to know how much license costs the candidates cause. Therefore, they use another indicator-based analysis technique "License costs" (c.f. [LL12], p. 187) to get information about the license costs. These costs are displayed as additional column in the “Technical Component Overview Viewpoint”. It appears that the license costs of TC1, TC2 and TC4 are very high. Figure 3 illustrates the analysis results. They also document this evaluation.

Name...	License Costs
TC1	100.000
TC2	80.000
TC3	85.000
TC4	0
TC5	10.000
TC6	30.000

Fig. 3 Technical Component Overview Viewpoint (bad state of health is highlighted)

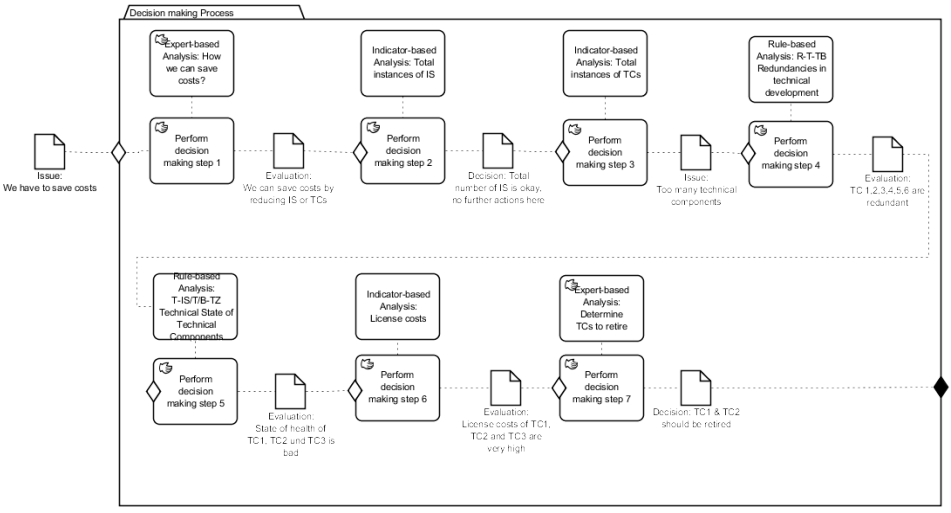


Fig. 4 CMMN model illustrating the scenario

After getting this information the stakeholders feel equal to take an adequate decision. They put together all the information they have and discuss. Result of this discussion is a decision that TC1 and TC2 should be retired, because they have a bad state of health and also cause much license costs. A decision is created to document the design decision. The responsibilities of this decision correspond to the participated stakeholders. By using the description field, which is available for all annotation type, they are able to document the rational, too. Next, they have to analyze the impacts of the technical component candidates on information systems. Here we assume that their suggestion is realizable. Therefore, Joe, the technology expert, agrees, that the functionality of TC1 and TC2 can also be provided by the other technical components. Figure 4 illustrates the steps, used analysis techniques and the steps' result.

5 Conclusion

In this paper, we reflected on the basic structure of collaborative decision-making processes in EAM, interlinked the corresponding knowledge-intensive processes with the prevalent analysis techniques in the field [BMS09], and showed how the insights gained during decision-making can be documented using the annotation-based mechanism introduced in [JSZ15]. We applied adaptive case management to integrate the activity and the result-perspective and explained how annotations become part of the case file documenting a particular decision-making case. These annotations and their relationships to the analysis techniques, which they are based on, provide a reasonable rationalization of the decision making process.

Future works in this field can further the presented approach. In the presented setup, the stakeholders are required to decide on the used viewpoints as well as on the analysis techniques to be applied based on their expert knowledge. The experts have to consider existing EAM goals to choose the appropriate techniques, rules and KPIs. Further, issues regarding particular EA artefacts are likely to call for certain analysis techniques – again the selection of the technique has to be based on experts' knowledge and experience. As we expect that with the advent of new paradigms, like cloud computing and the Internet of Things, new analysis techniques will emerge as well, experts are constantly challenged to expand their knowledge. We would assume that stakeholder profiles and recommendations based thereon, like discussed by Govedarski et al. in [GHS15] can also be applied to analysis techniques to facilitate the selection of the appropriate technique given the current issue, stakeholder, and EAM goals.

References

[BMS09] Buckl, S.; Matthes, F.; Schweda, C.M.: Classifying Enterprise Architecture Analysis Approaches. In: The 2nd IFIP WG5.8 Workshop on Enterprise Interoperability (IWEI'2009), Valencia, Spain, pp. 66-79, 2009.

[Fi11] Fischer, L.: Taming the Unpredictable Real World Adaptive Case Management: Case Studies and Practical Guidance, Future Strategies, 2011.

[Fi14] Fischer, L.: Empowering Knowledge Workers, Future Strategies, 2014.

[GHS15] Govedarski, K.; Hauptman, C.; Schweda, C.M.: Bottom-up EA Management Governance using Recommender Systems. In: In: Lecture Notes in Informatics – Jahreskonferenz Digital Enterprise Computing, vol. P-244, Böblingen, Germany, Köllen Verlag, 2015.

[Ha13] Hanschke, I.: Strategisches Management der IT-Landschaft: Ein praktischer Leitfaden für das Enterprise Architecture Management (3rd Edition), Hanser Verlag, München, 2013.

[Ha14] Hauder, M.; Münch, D.; Michel, F.; Utz, A.; Matthes, F.: Examining Adaptive Case Management to Support Processes for Enterprise Architecture Management. In: International Enterprise Distributed Object Conference Workshops and Demonstrations (EDOCW), Ulm, Germany, pp. 23-32, 2014.

[HPM14] Hauder, M.; Pigat, S.; Matthes, F.: Research Challenges in Adaptive Case Management: A Literature Review. In: International Enterprise Distributed Object Conference Workshops and Demonstrations (EDOCW), Ulm, Germany, pp. 98-107, 2014.

[HK15] Hamm, T.; Kehrer, S.: Goal-oriented Decision Support in Collaborative Enterprise Architecture. In: Lecture Notes in Informatics – Jahreskonferenz Digital Enterprise Computing, vol. P-244, Böblingen, Germany, Köllen Verlag, 2015.

[IT07] IT Governance Institute: COBIT 4.1, 2007.

[JE07] Johnson, P.; Ekstedt, M.: Enterprise Architecture – Models and Analyses for Information Systems Decision Making, Studentlitteratur, 2007.

[JS14] Jugel, D.; Schweda, C.M.: Interactive Functions of a Cockpit for Enterprise Architecture Planning. In: International Enterprise Distributed Object Conference Workshops and Demonstrations (EDOCW), Ulm, Germany, pp. 33-40, 2014.

[Ju15] Jugel, D. et al.: Providing EA Decision Support for Stakeholders by Automated Analyses. In: Lecture Notes in Informatics – Jahreskonferenz Digital Enterprise Computing, vol. P-244, Böblingen, Germany, Köllen Verlag, 2015.

[JSZ15] Jugel, D.; Schweda, C.M.; Zimmermann, A.: Modeling Decisions for Collaborative Enterprise Architecture Engineering. In: Persson, A.; Stirna, J. (Eds.): CAiSE 2015 Workshops, LNBIP 215, Springer, 2015.

[Ke08] Keim, D.A. et al.: Visual Analytics: Definition, Process, and Challenges. In: Information Visualization, Lecture Notes in Computer Science, vol. 4950, Springer Verlag, Heidelberg, pp. 154-175, 2008.

[LL12] Legner, C.; Löhe, J.: Embedding EAM into operation and monitoring. In: Ahlemann, F. et al. (eds.): Strategic Enterprise Architecture Management: Challenges, Best Practices and Future Developments. Springer Heidelberg, pp. 169-200, 2012.

[Ma08] Matthes F. et al.: Enterprise Architecture Tool Survey 2008, Technical Report, Technical University Munich, Chair for Informatics 19, München, Germany, 2008.

[Ma10] Mayring, P.: Qualitative Inhaltsanalyse (11th edition), Beltz, 2010.

[Ma11] Matthes, F.: EAM KPI Catalog v.1.0, Technical Report, Technical University Munich, Chair for Informatics 19, München, Germany, 2011.

[NSV14] Naranjo, D.; Sanchez, M.; Villalobos, J.: Towards a Unified and Modular Approach for Visual Analysis of Enterprise Models. In: International Enterprise Distributed Object Computing Conference Workshops and Demonstrations (EDOCW), Ulm, Germany, pp. 77-86, 2014.

[Om14a] Object Management Group: Case Management Modeling Notation 1.0, 2014.

[Om14b] Object Management Group: Decision Model and Notation 1.0 - Beta 1, 2014.

[PDP14] Plataniotis, G; De Kinderen, S.; Proper, H.A.: EA Anamnesis: An Approach for Decision Making Analysis in Enterprise Architecture. In: International Journal of Information Systems Modeling and Design. Vol. 4 (1), 2014, pp. 75-95.

[Ra14] Ramos, A. et al.: Automated Enterprise-Level Analysis of ArchiMate Models. In: Proceedings of the 15th International Conference, BPMDS 2014, 19th International Conference, EMMSAD 2014, held at CAISE 2014, Thessaloniki, Greece, pp. 439- 453, 2014.

[RZM14] Roth, S.; Zec, M.; Matthes, F.: Enterprise Architecture Visualization Tool Survey 2014, Technical Report, Technical University Munich, Chair for Informatics 19, München, Germany, 2014.

[Sa14] Sandkuhl, K.; Stirna, J.; Persson, A.; Wißotzki, M.: Enterprise Modeling Tackling Business Challenges with the 4EM Method, Springer Berlin Heidelberg, 2014.

[Sw10] Swenson, K. D.: Mastering the Unpredictable: How adaptive case management will revolutionize the way that knowledge workers get things done, Meghan-Kiffer Press, 2010.

[Sw13] Swenson, K. D.: State of the Art In Case Management, White Paper Fujitsu, 2013.

[TOG12] The Open Group: ArchiMate 2.0 Specification, 2012.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Daten-getriebene Unternehmensarchitekturen im E-Commerce für das präventive Retourenmanagement

Michael Möhring[1], Rainer Schmidt[2]

Abstract: Online-Händler sind Nutznießer der Digitalisierung aber für sie entstehen auch neue Risiken in der digitalen Gesellschaft. So hat die anonyme Käufer-Verkäufer-Beziehung sowie die fehlende reale Einschätzung des Produkts vor dem Kauf dazu geführt, dass Rechte wie das Produkte zurückzusenden viel häufiger ausgeübt werden als früher im stationären Einzelhandel. Diese Retouren stellen für viele Unternehmen ein erhebliches finanzielles Risiko dar. Zu seiner Bewältigung ist eine grundsätzliche Neustrukturierung der Unternehmensarchitekturen für Online-Händler notwendig. Eine Daten-getriebene Architektur löst die Entscheidungsvorgänge aus den Geschäftsprozessen heraus und ermöglicht so ihre separate Modellierung und Umsetzung. Hierdurch wird eine deutliche Beschleunigung von Entscheidung ermöglicht, die wiederum eine Unterstützung für das präventive Retourenmanagement schafft.

Keywords: Retourenmanagement, Retourenvermeidung, Präventives Retourenmanagement, E-Commerce. Daten-getriebene Unternehmensarchitekturen, EAM, product returns management

1 Einleitung

Neben etablierten Onlinehändlern wie Amazon oder Zalando wagen auch immer mehr traditionelle Einzelhändler wie bspw. Peek & Cloppenburg den Weg in den Internethandel. Im Gegensatz zum stationären Geschäft müssen dabei einige Geschäftsprozesse (wie bspw. Logistik-, Verkaufs- oder Zahlungsprozesse) grundlegend neu implementiert werden [Koll13]. Diese veränderten betrieblichen Funktionen sowie eine notwendige Echtzeitreaktion auf Ereignisse eines Webshops der 24h geöffnet hat, verlangt nach angepassten IT-Architekturen. Neben dem reinen Verkaufsprozess rücken auch Prozesse der Warenrückgabe in den Fokus, da der Konsument die Ware vor Kauf nicht physisch begutachten kann [ShCS10]. Retouren aufgrund von Nichtgefallen sind in einigen Branchen wie dem Modehandel mit mehr als 50% vorzufinden [PSWW13]. Neben Händlern aus dem klassischen Filialgeschäft haben auch die etablierten Onlinehändler mit Retouren zu kämpfen. Die Abwicklung von Retouren ist teuer und belastet die Umwelt [Lütg14]. Nach aktuellen Untersuchungen fallen für Onlinehändler durchschnittlich 15,18 Euro Gesamtkosten für eine Retoure an [Asde15]. Bisherige Forschungen zur Retourenvermeidung haben einen Rahmen zur Retourenprävention erarbeitet, Grundlagen der Nutzung von modernen Datenanalysetechniken entwickelt und das Retourenverhalten untersucht [WMKS14] [MWSK13] [MSKW13]. Wie IT-Architekturen jedoch ein Fundament zur IT-Unterstützung der Retourenvermeidung der sehr Daten- und Analyse-intensiven Prozesse bieten kann, ist nur sehr unzureichend untersucht. Der vorliegende Aufsatz zeigt erste Möglichkeiten der Gestaltung von Unternehmensarchitekturen für Onlineshops

[1] Friedrich-Schiller-Universität Jena, Lehrstuhl für ABWL und Marketing, Carl-Zeiß-Straße 3, 07743 Jena, michael.moehring@uni-jena.de
[2] Hochschule München, Fakultät für Informatik und Mathematik, Lothstraße 34, 80335 München, Rainer.Schmidt@hm.edu

zur datengetriebenen Retourenvermeidung auf. Der Aufsatz gliedert sich in nachfolgende Bereiche: Zunächst werden Grundlagen im Bereich des präventiven Retourenmanagements und den bisherigen Stand von Unternehmensarchitekturen von Onlineshops gelegt. Dann erfolgt die Konzeption einer Daten-getriebenen Architektur zur Unterstützung des präventiven Retourenmanagements. Abschließend werden ein Ausblick und eine kurze Aufstellung verwandter Forschungsarbeiten aufgezeigt.

2 Retouren als Herausforderung im E-Commerce

Konsumenten schicken online gekaufte Waren häufig zurück, weil diese nicht ihren Vorstellungen entsprechen [PSWW13]. Anders als im regionalen Einzelhandel können die Konsumenten den Artikel vor Kauf nicht an- bzw. ausprobieren [ShCS10]. Daher ist eine Retourenmöglichkeit im Onlinehandel unerlässlich und von vielen Onlinehändlern wie etwa der Otto Group als Bestandteil ihrer Geschäftsstrategie zu sehen [Otto15]. Dennoch belasten Retourenkosten das Geschäft vieler Onlinehändler [Asde15] [Meie00].

Nach Walsh und Möhring fallen für Onlinehändler vor allem Kosten des Handlings (Abwicklung der Retouren) und Kosten der Retourenvermeidung an [WaMö15]. Um derartige Kosten zu vermeiden und ebenfalls nachhaltige Kundenbeziehungen zu pflegen, ist ein präventives Retourenmanagement unerlässlich.

Das präventive Retourenmanagement versucht nach Walsh und Möhring Retouren vor und nach der Bestellung durch den Konsumenten zu vermeiden [WaMö15]. Aktuelle Forschungen im deutschsprachigen Raum haben einen Rahmen für das präventive Retourenmanagement entwickelt [WMKS14] [MWSU15] [MWSK13]:

Monetäre Instrumente	**Ablauforientierte Instrumente**	**Konsumenten-basierte Instrumente**
Geld-zurück-Garantien, Restocking-Fee, Gutscheine/Rabatte bei Nichtretoure, etc.	Sichere Verpackungen, Durchlaufzeitoptimierung, Keine Beilegung von Retourenscheinen, etc.	Virtual-Try On, Produkt-bewertungen, Avatare, Größentabellen, etc.

Abb. 1: Instrumente des präventiven Retourenmanagements [WMKS14]

Nach Walsh et al. [WMKS14] versuchen Konsumenten-basierte Instrumente Retouren vor dem Kauf vor allem durch bessere Informationen der Produkteigenschaften und der Passgenauigkeit zu vermeiden. Beispielhafte Instrumente hierbei sind etwa Produktbewertungen, Avatare, Virtual Try-On oder Größentabellen (vgl. Abb. 1). Monetäre Instrumenten geben den Kunden einen Anreiz (oder ein Hemmnis) Artikel nicht zu retournieren [WMKS14]. Geld-zurück-Garantien, wie sie beispielsweise bei Zalando mit 100-Tagen anzufinden sind, Wiedereinlagerungsgebühren (sog. Restocking-Fees) oder etwa Gutscheine bei Nichtretoure sollen Rücksendungen vermeiden (vgl. Abb. 1). Ablauforientierte Instrumente vermeiden Retouren nach der eigentlichen Kundenbestellung im Onlineshop [WMKS14]. Dazu zählen etwa sichere Verpackungen, Durchlaufzeitoptimierungen oder Instrumente welche die Schikanekosten (sog. Hassle-costs) (etwa keine Retourenscheine beilegen) erhöhen (vgl. Abb. 1).

Weiterhin wurde in verschiedenen Studien das Retourenverhalten deutscher Konsumenten untersucht (vgl. bspw. [PSWW13] [MWSU15]) und Möglichkeiten von Big Data und Predictive Analytics zur Retourenvermeidung dargestellt (vgl. bspw. [WaMö14] [MWSK13] [MSKW13].

Grundlegend sollten Maßnahmen zur Retourenvermeidung kunden- und produktspezifisch schon vor der Kaufphase implementiert werden [WaMö14] [MWSK13]. Die Implementierung dieser Maßnahmen bedarf einen hohen Datenanalyseaufwand, welche durch die IT-Architektur des Online-Shops unterstützt werden muss.

3 IT-Architekturen von Online-Shops

Vor allem Geschäftsziele wie die Retourenreduktion und Margenerhöhung sind neben IT-Zielen wie der durchgängige IT-Betrieb mit gutem Antwortzeitverhalten der IT-Systemen wesentliche Ziele von Onlinehändlern mit Bezug auf das präventive Retourenmanagement auf IT Architekturen.

Im Allgemeinen werden Architekturprinzipien von Geschäfts- und IT-Zielen beeinflusst und münden in einem Design, Repräsentationsregeln, Guidelines und Evaluationskriterien [Stel10]. Grundlage für die Architekturentwicklung waren bisher vorwiegend service-orientierte Konzepte [Chaf07]. In jüngster Zeit wird aber zunehmend auch Cloud-Computing [MeGr09] in die Architekturüberlegungen einbezogen [KaMa11]. Zur methodischen Unterstützung empfiehlt sich die Verwendung von Enterprise Architecture Management [Lank05].

Onlineshops sind von einer sehr heterogenen IT-Infrastruktur durch Einsatz verschiedener IT-Architekturkomponenten ähnlich wie der Einzelhandel (vgl. bspw. [SMZH14] [Koll13] [Merz01]) geprägt. Dabei verwenden Onlinehändler je nach Historie und Ausprägung ein Warenwirtschaftssystem (WWS) bzw. Enterprise-Ressource-Planning (ERP)-System, welches die Transaktionen verwaltet [MBKP12], siehe auch Abbildung 2. Darauf aufbauend wird eine Online-Shop-Software (bspw. OSCommerce, Intershop, etc.) betrieben, welche Transaktionen mit dem Endkunden ermöglichen. Dabei erfolgt stetig oder in gewissen Zeitabständen ein Abgleich der gespeicherten Stamm- und Bewegungsdaten zwischen dem WWS- bzw. ERP-System und der Online-Shop-Software. Neben diesen Transaktionssystemen bestehen je nach Ausrichtung noch Managementinformationssystemen wie ein Data-Warehouse mit Business-Intelligence-Software [KeBL13]. Diese sind jedoch i.d.R. nur sehr unzureichend mit den Transaktionssystemen (wie dem Onlineshop) integriert, um diese im operativen Betrieb nicht mit zu langen Lese- und Schreibzyklen zu belasten. Weiterhin können noch Dritt-Systeme und Schnittstellen zu externen Partnern (wie Logistik-, Finanzdienstleister) bestehen. Zusammengefasst ist eine sehr heterogene Architektur vorzufinden, welche auf die schnelle Ausführung von Transaktionen ausgelegt ist, nicht aber die Durchführung von Analysen und das Fällen von Entscheidungen in Echtzeit. Entscheidungen in Echtzeit während der Online-Shop-Besucher auf der Seite ist, kann demnach bisher nur unzureichend oder gar nicht technisch unterstützt werden. Vor allem aber zur Implementierung von Strategien des präventiven Retourenmanagements (ggf. Echtzeitmanipulation von Größenauswahlen und Produkten) ist eine gute Analysearchitektur unerlässlich [MWSK13].

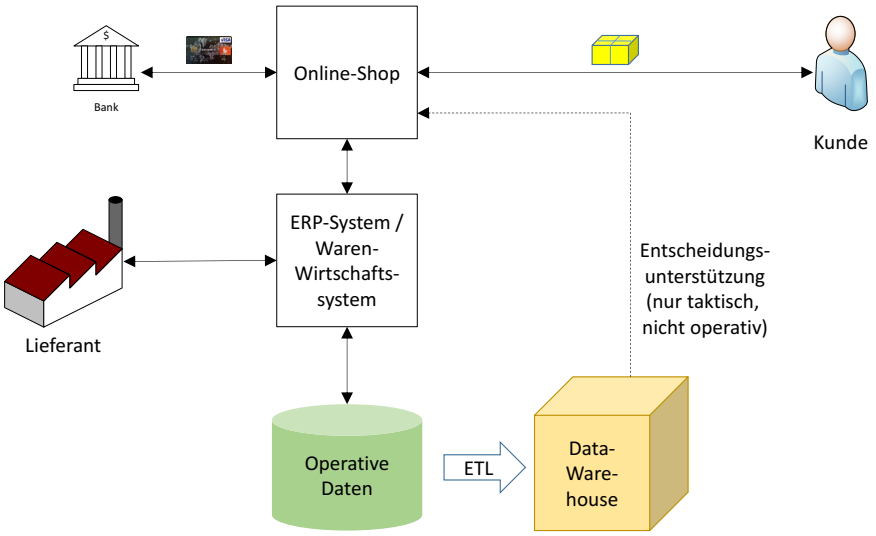


Abb. 2: Typische Architektur von Online-Shops

Wichtige Bestandteile des präventiven Retourenmanagements erfordern die Durchführung von Analysen und die Entscheidungsfindung meist in Echtzeit (bzw. ohne große zeitliche Verzögerungen). Die oben beschriebene Architektur ist hierzu aus zwei Gründen nicht in der Lage.

Durch die Verwendung eines Data-Warehouses und der Übertragung von Daten in dieses über einen Extract-Transform-Load-Prozess [VaSS02] entsteht eine erhebliche Latenz. Typisch ist eine stapelartige (batch-oriented) Arbeitsweise. Dabei wird eine Reihe von Datenzwischenprodukten erzeugt, die für unterschiedliche Analysezielstellungen verwendet werden kann. Wegen des zeitlichen Aufwandes für die Erstellung der Datenzwischenprodukte weisen die etablierten Ansätze eine erhebliche Entscheidungslatenz auf, d.h. es liegt ein erheblicher Zeitraum zwischen der prinzipiellen Verfügbarkeit der Informationen und dem Vorhandensein der Analyse. Diese hohe Latenz ist bei strategischen Entscheidungsprozessen meist nicht besonders problematisch, für die Entscheidungsunterstützung auf operativer Ebene hingegen, ist die Latenzzeit zu groß. Beispielsweise geht es bei der Fehleranalyse darum, innerhalb kürzester Zeit Fehlerursachen zu identifizieren. So können mitunter Tage vergehen bis die Informationen für die Entscheidungsfindung bereitstehen. Für die hier angestrebte Entscheidungsfindung in Echtzeit ist dieser Ansatz somit unbrauchbar

Das zweite Problem besteht in der Einbettung der Entscheidungsfindung in die Anwendungssysteme. Ähnliche wie bei einem Datenbanksystem sollte die Funktionalität zur Unterstützung der Entscheidung getrennt von den übrigen Funktionalitäten des Gesamtsystems gehalten werden. Nur so kann ein schnelle Anpassung und Erweiterung der Entscheidungsunterstützung erreicht werden. Bei dem oben dargestellten Ansatz ist die Entscheidungsunterstützende Funktionalität jedoch stark über das Gesamtsystem verteilt.

4 Datengetriebene Unternehmensarchitektur für die Unterstützung des präventivem Retourenmanagements

Um eine präventive Vermeidung von Retouren vor dem Kauf zu ermöglichen, sind vor allem kunden- und produktspezifische Analysen nötig [MWSK13]. Zum einen sollten Retourenmuster und Präferenzen des Kunden (bspw. durchschnittlich behaltene Konfektionsgröße) ermittelt werden, um beispielsweise Artikel vorzuselektieren oder Größen auszublenden [MWSK13]. Zum anderen sind Retourenmuster von Produkten zu ermitteln (bspw. welche Abweichungen treten auf und was ist deren Echtgröße bzw. Spezifikation) [MWSK13].

Weiterhin sollten standardisierte Vorgehensweisen je nach Kunden- und Produktgruppe implementiert werden. Aus bisherigen Studien [MWSU15] ist bekannt, dass weibliche Konsumenten mehr Mode retournieren als dies männliche tun. Neben dem Geschlecht existieren auch etwa altersbedingte Unterschiede beim Retournieren. Jüngere Konsumenten retournieren öfter Mode als dies ältere tun. Neben diesen soziodemografischen Merkmalen ist auch bspw. bekannt, dass das Retourenverhalten von anderen Produktkategorien (wie bspw. Elektronikartikel) das Retournieren von Modeartikeln beeinflusst [MWSU15]. Demnach sollte auch das Verhalten verschiedener Produktkategorien erfasst und ausgewertet sowie präventiv genutzt werden. Um Retourenmuster (bspw. Produktspezifikationsabweichungen wie Größe, Farbe, etc.) zu erkennen, können auch Konsumentenmeinungen genutzt werden [MWSK13] [WaMö14].

Dabei kann zum einen auf eigene Produktbewertungen im Onlineshop zurückgegriffen werden, oder aber Produktbewertungen auf größeren Plattformen (wie ciao.de etc.) oder Händlern (wie Amazon.de). Eine Produktidentifizierung ist dabei über die EAN, ASIN, UPC oder Kontextinformationen der Produktkategorie möglich [MWSK13]. Neben Daten aus Produktbewertungsportalen, kann auch auf Daten aus sozialen Netzwerken wie Facebook oder Twitter zurückgegriffen werden [MWSK13]. Hierbei besteht jedoch die Herausforderung Produkte richtig zu identifizieren [MWSK13].

Die unterschiedlichen Datenquellen und Datenverarbeitungsorientierungen sind in der nachfolgenden Tabelle dargestellt:

	Attribute	**Datenquellen**	**Orientierung der Datenverarbeitung**
Konsumenten-daten	Alter	ERP, WWS, Onlineshop, etc.	Transaktion
	Geschlecht	ERP, WWS, Onlineshop, etc.	Transaktion
	Kundenwert (CLV)	ERP, BI-System	Analyse
Produktdaten	Produktstammdaten	ERP, WWS, Onlineshop	Transaktion
	Produktbewertungen	Onlineshop,Produktbewertungs-portale, soziale Netzwerke	Analyse
Retouren-management-daten	Handlingkosten	ERP, WWS	Analyse
	Kundenretourenverhalten	ERP, WWS, etc.	Analyse
	Kunden-Retourencluster	ERP, BI-System	Analyse
	Produkt-Retourencluster	ERP, BI-System	Analyse
	Produkt-Retourengründe	ERP, WWS	Analyse
	Produkt-Retourenhäufigkeiten	ERP, WWS, BI-System	Analyse

Tab. 1: Datenquellen und -verarbeitungsorientierungen (in Anlehnung und erweitert nach [MWSK13])

Basierend auf den Datenquellen und Datenverarbeitungsorientierungen (vgl. Tab. 1) wird deutlich, dass vor allem zur Analyse des Retourenverhaltens (wie den Retourenmanagementdaten) eine Analyseorientierung notwendig ist [MWSK13].

Grundlage der hier vorgestellten Architektur ist die Trennung von Prozessen und Entscheidungs-Services. Betriebliche Entscheidungen werden als eigenständige Entitäten separat vom Prozessmodell verwaltet. Die Entscheidungs-Services enthalten aber nicht nur das Entscheidungsverfahren, sondern auch die Darstellung der Entscheidungsgrundlage sowie die entscheidungsrelevanten Daten. Hierdurch kann bei Änderungen der Entscheidungsgrundlage automatisiert reagiert werden.

Durch die separaten Entscheidungs-Services (siehe Abbildung 3) werden die Wiederverwendung, die Korrektheitsprüfung und die Geschwindigkeit von Veränderungen der Entscheidungsverfahren erhöht. Durch die eigenständige Darstellung können Entscheidungen in mehreren Prozessen standardisiert und wiederverwendet werden. Im Gegensatz hierzu ist die tiefe Einbettung von Entscheidungen in Geschäftsprozessmodelle wie beispielsweise nach dem BPMN Standard [ChTr12] für die Wiederverwendung höchst problematisch.

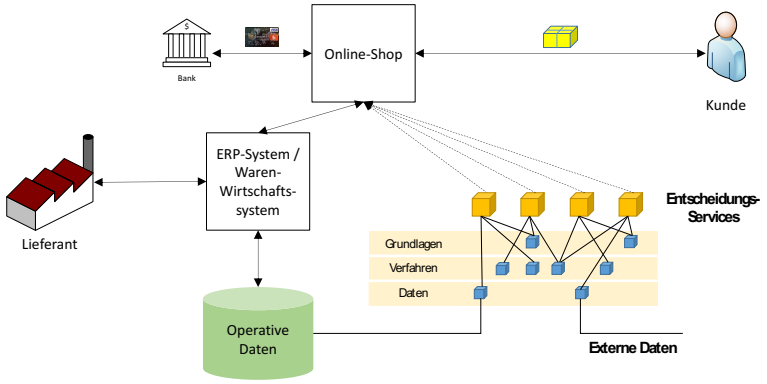


Abb. 3: Separate Darstellung des Entscheidungsmodels

Auch die Korrektheitsüberprüfung profitiert von der separaten Darstellung. Die als eigenständige aber abgeschlossene Entität dargestellten Entscheidungen können auf Ihre Korrektheit und Konsistenz geprüft werden. Existierende Ansätze wie beispielsweise Business Rules [TeOt06] sind eine Sammlung von einzelnen Regeln, deren Korrektheit bezüglich einer konkreten Entscheidung in einem Geschäftsprozess nur sehr aufwendig, wenn überhaupt, überprüfbar ist.

Die Erfassung von Entscheidungsgrundlagen, -verfahren und –daten ermöglicht zudem schneller und umfassender auf Veränderungen zu reagieren. Bisherige Entscheidungsverfahren stützten sich nur auf die grundlegenden Daten, sahen aber keine Reaktion auf Veränderungen in der Entscheidungsgrundlage und dem Entscheidungsverfahren vor.

Durch die separaten Entscheidungs-Services (siehe Abbildung 3) kann die Korrektheit, Wiederverwendung und Agilität des Prozesses erhöht werden. Die Entscheidungsservices greifen auf interne und externe Services zurück. Auf diese Weise können auch Änderungen der Entscheidungsgrundlagen und des Entscheidungsverfahrens berücksichtigt werden. Diese Services werden aus internen und externen Service-Kandidaten mit Hilfe eines Reputationsmodelles ausgesucht welches Sicherheit, Datenqualität und die unterstützten Meta-Services.

5 Umsetzung mit Hilfe verteilter Verarbeitungsmechanismen

Aktuelle Forschungen zum Einsatz von Unternehmensarchitekturen beschäftigen sich mit Komponenten von Big Data, um so zum einen Unternehmensarchitekturen besser beurteilen zu können, als auch daten-intensive Geschäftsprozesse zu ermöglichen [SMZW14] [SMZH14] [ZPZF13] [Ferg12] [ScKü15]. So sind beispielsweise für den Bekleidungseinzelhandel bereits derartige Architekturen entwickelt worden [SMZH14].

Grundlage für datengetriebene Architekturen sind verteilte Verarbeitungsmechanismen welche vielfach unter dem Begriff Big Data [BuCM10] zusammengefasst werden. Sie unterscheiden sich deutlich von den oben beschriebenen Konzepten. Ein erster wichtiger Unterschied liegt in der Fähigkeit auch semi- und unstrukturierte sowie unvollständige Daten zu bearbeiten. Semi-strukturierte Daten sind Daten ohne vordefiniertes Schema, bei denen aber dieses Schema zur Laufzeit bestimmt werden kann. Der hierfür notwendige Aufwand schränkte bisher ihre Nutzung ein. Nach der Schemaerkennung sind semi-strukturierte Daten nutzbar wie strukturierte Daten. Unstrukturierte Daten sind Daten wie beispielsweise Kundenäußerungen in Social Media Systemen [WaKH11] [MSHH14]. Kennzeichnend für diese Art der Daten ist, dass sie keinem einheitlichen Schema unterliegen. Es liegt dabei eine sogenannte semantische Heterogenität vor [MSHH14]. Hierunter ist die unterschiedliche Interpretation von Begriffen in unterschiedlichen Kontexten zu verstehen. Daher kann hier die Auswertung nur über statistische Verfahren oder maschinelle Lernverfahren erfolgen, welches den Verarbeitungsaufwand gegen über semi-strukturierten Daten weiter erhöht [MWSK13].

Ein weiteres Unterscheidungsmerkmal des zu entwickelnden Software-Systems ist der Typ der unterstützen Entscheidungs-Services [SMZW14]. Bisherige Systeme sind größtenteils nur zu berichtenden, d.h. vergangenheitsbezogenen Analysen fähig. Das zu entwickelnde Software-System wird auch vorhersagende und empfehlende Analysen ermöglichen. Für die Durchführung von vorhersagenden oder empfehlenden Analysen ist es notwendig, eine möglichst breite statistische Grundlage einzubeziehen. Um größere Volumina auch semi- und unstrukturierter Daten in kurzer Zeit auszuwerten und zur Grundlage von Vorhersagen und Entscheidungen zu machen setzt das zu entwickelnde System hoch verteilte Verfahren ein. Die vorhandenen Werkzeuge können auch an Hand ihrer Latenz verglichen werden. Hierunter wird der zeitliche Verzug bei der Erstellung von Analysen verstanden. Das Bestreben der etablierten Systeme ist es von monatlichen und wöchentlichen Auswertungen hin zu tagesaktuellen, stündlichen oder gar echtzeitnahen Analysen zu gelangen. Besonders hierdurch kann die Findung nicht nur strategischer, sondern auch operativer Entscheidungen unterstützt werden. Die etablierten Verfahren sind für strategische Entscheidungen einsetzbar, für operative Entscheidungen benötigen diese meist zu lange. Insbesondere um schnell auf Fehlfunktionen zu reagieren oder die Ressourcenverteilung anzupassen ist es notwendig, in Sekundenbruchteilen Entscheidungen zu treffen.

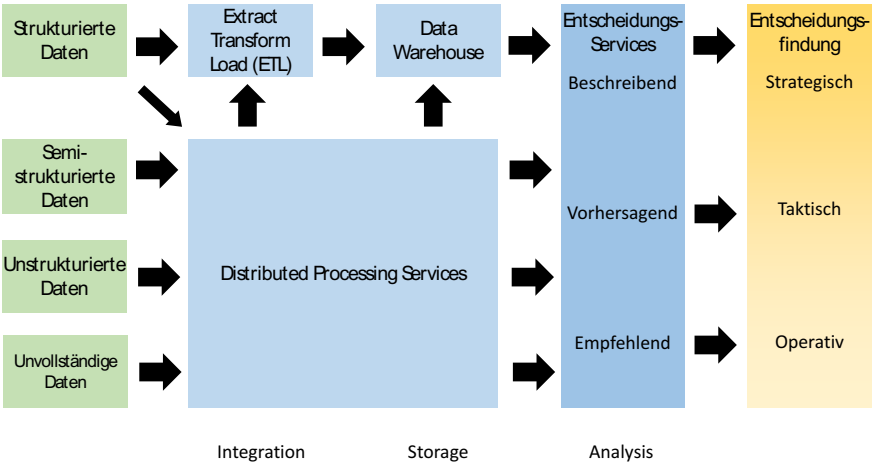


Abb. 4: Erweiterung der Entscheidungsservices und –findung in Anlehnung an [SMZW14]

Für eine analyse-orientierte Datenverarbeitung, welche in Echtzeit während des Kundenbesuches erfolgt, sind derzeitige Architekturen bei vielen Onlineanbietern nur wenig ausgelegt (vergleiche Kapitel davor). Weiterhin variieren Besucherströme je nach Tageszeit [Brei07]. Daher ist nicht zu jedem Zeitpunkt die gleiche Rechenkapazität nötig. Neben den Verarbeitungskapazitäten ist für eine analyse-orientierte Datenverarbeitung auch ein vertieftes Wissen über mathematisch, statistische Algorithmen (sog. Data-Scientists) und entsprechende Business-Intelligence- und Data-Miningsoftware nötig [Pati11], [Kotu14], [MüLe13].

Daher können vor allem neu aufkommende Analytics-Services von Cloud-Anbietern von Amazon oder Microsoft genutzt werden. Microsoft bietet mit Azure ML ein teilweise auf R-Project basierendes skalierbares Produkt an, mit dem Unternehmen Retourenmuster ermitteln können. Ähnliches ist auch bei anderen Anbieter erhältlich (siehe Tabelle 2).

Eine Auswahl an Analysetools für den vorgestellten Business Case ist in der nachfolgenden Tabelle dargestellt. Es wird weiterhin dabei deutlich, welche Lösungen entgeltpflichtig sind und bei welcher Lösung Anforderungen an eine einfache Skalierbarkeit sowie Cloud-Lösung geben sind (siehe Tabelle 2).

Basierend auf einer Nutzwertanalyse mit den in der Tabelle 2 gezeigten Elementen je Analyseumgebung und Erfahrungen mit den Anbietern (wie Integrationsfähigkeit) kann eine Architekturentscheidung über die jeweiligen Cloud-Services erfolgen [SMZW14].

Neben dieser reinen, starren Cloud-Anwendung besteht auch die Möglichkeit selbst derartige Services auf einer Private- oder Public-Cloud-Umgebung zu implementieren. So können beispielsweise mit der Analysesoftware RapidMiner erstellte Analyseprozesse als Web-Service über die serverseitige Plattform RapidAnalytics bereitgestellt werden.

Analyseumgebung	Kostenaspekte	Cloud-Service	Einfache Skalierbarkeit
Microsoft Azure ML [Barg14]	Entgeltpflichtig	Ja	Ja
Amazon Machine Learning [Aml15]	Entgeltpflichtig	Ja	Ja
Google Prediction API [Go15]	Entgeltpflichtig	Ja	Ja
RapidMiner Cloud [Rmc15]	Entgeltpflichtig	Ja	Ja
RapidMiner [Kotu14]	Entgeltfreie sowie entgeltpflichtige Variante verfügbar	Nein	Nein
R-Project [Lant13]	Entgeltfrei	Nein	Nein
IBM SPSS Modeller [Ibm15]	Entgeltpflichtig	Nein	Nein
SAS 9.4 [Sas15]	Entgeltpflichtig	Nein	Nein
Microsoft SQL Server Business Intelligence [BaSa14]	Entgeltpflichtig	Nein	Nein
SAS Cloud Analytics [Sas15]	Entgeltpflichtig	Ja	Ja

Tab. 2: Auswahl und Klassifikation von Analysetools

6 Verwandte Arbeiten

Die modellhafte Darstellung von Entscheidungen wurde bereits in der Vergangenheit durch eine Reihe von Verfahren repräsentiert. Die Predictive Model Markup Language (PMML) [GZLW09], [Pech09] ermöglicht die standardisierte Darstellung von vorhersagenden Modellen. Sie kann zur Darstellung von Wissensmodellen als Bestandteil von Entscheidungspatterns herangezogen werden. Allerdings fehlt jede Integration in Geschäftsprozesse. Durch den SBVR Standard, Semantics of Business Vocabulary and Rules ist es möglich, Sachverhalte und Regeln aus dem Business Bereich semantisch reichhaltig darzustellen [TeOt06]. Der OMG-Standard Decision Model And Notation (DMN) [Omg15] ermöglich die grafische Repräsentation von Entscheidungsvorgängen in Geschäftsprozessen, gibt aber keine tiefere methodische Unterstützung. Der Geschäftsprozessmodellierungsstandard Business Process Modelling Notation [ChTr12] enthält Konstrukte zur Darstellung von Entscheidungen, diese eignen sich aber nicht für die Modellierung komplexer Entscheidungsalgorithmen.

Weiterhin beeinflussen auch neue Preisbildungsmechanismen wie „Dynamic Pricing“ [HiHS11] die Architekturwahl der Onlineshops, da hierbei sehr datenintensive Analyseprozesse durchgeführt werden müssen.

7 Fazit und Diskussion

E-Commerce Unternehmen müssen sich neben steigenden Umsätzen und wandelnden Konsumentenanforderungen (wie bspw. Same-Day-Delivery [Ec15]) vor allem auch auf steigende Retouren einstellen. Um Retouren daten-getrieben zu vermeiden, sind vor allem aufwendige Datenanalysen nötig. Der vorliegende Aufsatz baut auf Vorarbeiten im Bereich Business Intelligence und Big Data sowie des präventiven Retourenmanagements auf und stellt einen Ansatz zur Gestaltung von Unternehmensarchitekturen zur Unterstützung der Analyse und Durchführung von IT-gestützten Retourenvermeidungen vor.

Zentrale Idee ist dabei die Herauslösung der Entscheidungsabläufe aus den Geschäftsprozessen als separate Entscheidungsservices. Diese Entscheidungsservices werden auf der Basis untergeordneter Services bereitgestellt, die Daten, Entscheidungsmethoden und –grundlagen zur Verfügung stellen. Die den verwendeten Service-orientieren Ansatz können Änderungen leicht umgesetzt werden und auch externe Quellen für Daten, Entscheidungsmethoden und –grundlagen leicht genutzt werden. Durch diese Umgestaltung ist es möglich, automatische Entscheidungen deutlich schneller zu treffen als bisher und sogar in operative Prozesse einfließen zu lassen. Hierdurch wird ein wichtiger Schritt zum Daten-getriebenen Unternehmen ermöglicht.

Unser Ansatz erweitert die bisherige Forschung in vielfacher Hinsicht. Zukünftige Forschung kann von neuen Architekturmustern zur Unterstützung der Retourenvermeidung profitieren und somit die Umsetzbarkeit von Strategien des präventiven Retourenmanagements ermöglichen. Forschungen im Bereich Enterprise Architecture Management können ebenfalls von neuen Formen Daten-zentrierter Unternehmensarchitekutren im E-Commerce profitieren und diese mit vor allem bei kleineren und mittleren Onlineshops vorliegenden starren Transaktionsorientierung evaluieren und weiterentwickeln.

E-Commerce Manager können basierend auf unserem Ansatz Möglichkeiten zur Umsetzung von daten-getriebenen Retourenpräventionsstrategien evaluieren und mögliche Technologien und Architekturmuster selektieren. Mit dem vorliegenden Beitrag ist es möglich die bisherige Architektur eines Onlineshops durch Erweiterungen und Adaptionen zu einer mehr analyseorientierten Architektur zu transformieren, um so nachhaltig Wettbewerbsfähigkeit zu sichern.

Der vorliegende deskriptiv begründete Ansatz basiert auf Erkenntnissen der aktuellen Literatur und Branchenerfahrungen. Limitationen sind vor allem in Bereichen der empirischen Überprüfung des Ansatzes als auch der möglichen Branchen- und Größenunterschiede der Onlineshops zu finden.

Zukünftige Forschung sollte diesen Architekturvorschlag branchenübergreifend bei Onlineshops (bspw. mittels Case-Study-Research) evaluieren und weiterentwickeln. Unterschiede je nach Branche zur Retourenvermeidung (bspw. Elektronik und Modebranche) und der jeweiligen generischen Strategie (bspw. Premium-/Kostenführer) sollten näher betrachtet werden. Weiterhin sollte erforscht werden, wie Entscheidungsservices vorteilhaft gestaltet und software-technisch implementiert werden können.

Literaturverzeichnis

[Aml15] AML: Amazon Machine Learning – Make Data-Driven Decisions at Scale | AWS Official Blog, https://aws.amazon.com/de/blogs/aws/amazon-machine-learning-make-data-driven-decisions-at-scale/, Stand 15.05.2015.

[Asde15] Asdecker, B.: Statistiken Retouren Deutschland - Definition, http://www.retourenforschung.de/definition_statistiken-retouren-deutschland.html, Stand 15.05.2015.

[Barg14] Barga, R.: Predictive Analytics with Microsoft Azure Machine Learning. Apress, Berkeley, CA ; s.l., 2014.

[BaSa14] Bassan, A.S., Sarkar, D.: Mastering SQL server 2014 data mining: master selecting, applying, and deploying data mining models to build powerful predictive analysis frameworks, 2014.

[Brei07] Breidenbach, P.: E-Commerce bei Multi-Channel-Unternehmen. Analytische Betrachtung der Informations-und Kaufströme von Konsumenten und unternehmensstratehische Handlungsfelder für Anbieter, Berlin (Geographie der Kommunikation, Bd. 8, 2007.

[BuCM10] Bughin, J. et al.: Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. McKinsey Quarterly. 56, 2010.

[Chaf07] Chaffey, D.: E-business and E-commerce Management: Strategy, Implementation and Practice. Pearson Education, 2007.

[ChTr12] Chinosi, M., Trombetta, A.: BPMN: An introduction to the standard. Computer Standards & Interfaces. 34, 1, S. 124-134, 2012.

[CLSF05] Curbera, F. et al.: Web Services Platform Architecture: SOAP, WSDL, WS-Policy, WS-Addressing, WS-BPEL, WS-Reliable Messaging and More. Prentice Hall PTR, 2005.

[Ec15] EC: E-commerce: Same-day dreamers | The Economist, http://www.economist.com/news/business/21601556-online-firms-are-plunging-same-day-delivery-again-same-day-dreamers, Stand 15.05.2015.

[Ferg12] Ferguson, M.: Architecting a big data platform for analytics. A Whitepaper Prepared for IBM, 2012.

[Go15] Go: Prediction API — Google Cloud Platform, https://cloud.google.com/prediction/docs, Stand 15.05.2015.

[GZLW09] Guazzelli, A. et al.: PMML: An open standard for sharing models. The R Journal. 1, 1, S. 60-65, 2009.

[HeSt11] Heinrich, L.J., Stelzer, D.: Informationsmanagement: Grundlagen, Aufgaben, Methoden. Oldenbourg, R, 2011.

[HiHS11] Hinz, O. et al.: Price discrimination in e-commerce? An examination of dynamic pricing in name-your-own price markets. Mis quarterly. 35, 1, S. 81–98, 2011.

[Ibm15] IBM: IBM SPSS Software - Deutschland, http://www-01.ibm.com/software/de/analytics/spss/, Stand 15.05.2015.

[KaMa11] Kailer, D., Mandl, P.: Einsatz von Cloud Computing im E-Commerce. Praxis der Informationsverarbeitung und Kommunikation. 34, 1, S. 37-43, 2011.

[KeBL13] Kemper, H.-G. et al.: An Integrated Business Intelligence Framework. In: Rausch, P. et al. (eds.) Business Intelligence and Performance Management. S. 13–26, Springer London, 2013.

[Koll13] Kollmann, T.: E-Business: Grundlagen elektronischer Geschäftsprozesse in der Net Economy. Springer Gabler, Wiesbaden, 2013.

[Kotu14] Kotu, V.: Predictive analytics and data mining: concepts and practice with rapidminer. Elsevier, Waltham, MA, 2014.

[Lank05] Lankhorst, M.: Enterprise Architecture at Work: Modelling, Communication, and Analysis. Springer, 2005.

[Lant13] Lantz, B.: Machine learning with R: learn how to use R to apply powerful machine learning methods and gain an insight into real-world applications. Packt Publ, Birmingham, 2013.

[Lütg14] Lütge, G.: Online-Versand: Verflixte Retouren | ZEIT ONLINE, http://www.zeit.de/2014/15/retouren-onlinehandel-umweltbilanz, Stand 15.05.2015.

[MBKP12] Mertens, P. et al.: Grundzüge der Wirtschaftsinformatik. Springer Gabler, Berlin, 2012.

[MeGr09] Mell, P., Grance, T.: The NIST Definition of Cloud Computing, http://csrc.nist.gov/groups/SNS/cloud-computing/, Stand 15.05.2015.

[Meie00] Meier, S.: Retourenkosten belasten den Versandhandel, http://www.springerprofessional.de/retourenkosten-belasten-den-versandhandel/4832758.html, Stand 15.05.2015.

[Merz01] Merz, M.: E-Commerce und E-Business: Marktmodelle, Anwendungen und Technologien. dpunkt, Heidelberg, 2001.

[MSHH14] Möhring et al.: Neue Potenziale im Controlling durch die Verarbeitung von unstrukturierten Daten in Marketing und Vertrieb. Controlling Berater, 34, S. 229-245, 2014.

[MSKW13] Möhring, M. et al.: Big Data – neue Möglichkeiten im E-Commerce. Wirtschaftsinformatik und Management. 2013, 2, S. 48-56, 2013.

[MüLe13] Müller, R.M., Lenz, H.-J.: Business Intelligence. Springer, 2013.

[MWSK13] Möhring, M. et al.: Präventives Retourenmanagement im eCommerce. HMD Praxis der Wirtschaftsinformatik. 50, 5, S. 66-75, 2013.

[MWSU15] Möhring, M. et al.: Moderetouren im Deutschen Onlinehandel – Eine empirische Untersuchung. HMD Praxis der Wirtschaftsinformatik, 52 , 2, S. 257-266, 2015.

[Omg15] OMG: DMN 1.0, http://www.omg.org/spec/DMN/1.0/Beta1/, Stand 15.05.2015.

[Otto15] Otto: Retouren bei OTTO - Otto (GmbH & Co KG), https://www.otto.de/unternehmen/de/newsroom/dossiers/Retouren.php, Stand 15.05.2015.

[Pati11] Patil, D.J.: Building data science teams. O'Reilly Media, Inc., 2011.

[Pech09] Pechter, R.: What's PMML and What's New in PMML 4.0? ACM SIGKDD Explorations Newsletter. 11, 1, S. 19-25, 2009.

[PSWW13] Pur, S. et al.: Retourenmanagement im Online-Handel - Das Beste draus machen, http://www.ibi.de/files/Retourenmanagement-im-Online-Handel_-_Das-Beste-daraus-machen.pdf, Stand 15.05.2015.

[Rmc15] RMC: RapidMiner Cloud - RapidMiner, https://rapidminer.com/products/cloud/, Stand 15.05.2015.

[Sas15] SAS: Business Analytics und Business Intelligence Software | SAS, http://www.sas.com/de_de/home.html, Stand 15.05.2015.

[ScKü15] Schacht, S., Küller, P.: 2.2 Enterprise Architecture Management und Big Data. Praxishandbuch Big Data. 32, 2015.

[ShCS10] Shulman, J.D. et al.: Optimal reverse channel structure for consumer product returns. Marketing Science. 29, 6, S. 1071-1085, 2010.

[SMZH14] Schmidt, R. et al.: Datenzentrierte Unternehmensarchitekturen im Bekleidungseinzelhandel. INFORMATIK 2014 : Big Data - Komplexität meistern 44. Jahrestagung der Gesellschaft für lnformatik. LNI 232, S. 405-416, Stuttgart, 2014.

[SMZW14] Schmidt, R. et al.: Towards a Framework for Enterprise Architecture Analytics. Proceedings of the 18th IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW), Ulm / Germany, in press. , Ulm, Germany, 2014.

[Stel10] Stelzer, D.: Enterprise architecture principles: literature review and research directions. Service-Oriented Computing. ICSOC/ServiceWave 2009 Workshops. S. 12-21, Springer, 2010.

[TeOt06] Team, S., others: Semantics of Business Vocabulary and Rules (SBVR). Technical Report dtc/06–03–02, 2006.

[VaSS02] Vassiliadis, P. et al.: Conceptual modeling for ETL processes. Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP. S. 14-21 ACM, 2002.

[WaKH11] Walsh, G. et al.: Grundlagen des Web 2.0. Springer, Berlin Heidelberg, 2011.

[WaMö14] Walsh, G., Möhring, M.: Retourenvermeidung im E-Commerce—Kann Big Data helfen? Marketing Review St. Gallen. 31, 1, S. 68-78, 2014.

[WaMö15] Walsh, G., Möhring, M.: Definition Retourenmanagement; Gabler Wirtschaftslexikon, http://wirtschaftslexikon.gabler.de/Definition/retourenmanagement.html, Stand 15.05.2015.

[WMKS14] Walsh, G. et al.: PREVENTIVE PRODUCT RETURNS MANAGEMENT SYSTEMS-A REVIEW AND MODEL. Proceedings of the 21th European Conference on Information Systems (ECIS), Tel Aviv, Israel, 2014.

[ZPZF13] Zimmermann, A. et al.: Towards Service-Oriented Enterprise Architectures for Big Data Applications in the Cloud. Enterprise Distributed Object Computing Conference Workshops (EDOCW), 2013 17th IEEE International. S. 130-135, 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Current State of Enterprise Architecture Management in SME Utilities

Felix Timm[1], Matthias Wißotzki[2], Christina Köpp[3], Kurt Sandkuhl[4]

Abstract. In the last decades numerous developments and legal changes moved the utility industry towards a liberalized market. Utility enterprises have to stay competitive and reduce costs while managing more complex IT systems. The authors of this work see special demand for aligning business and IT for small and medium-sized enterprises (SME) in this industry and identify Enterprise Architecture Management (EAM) as the key for this issue. Therefore a survey was conducted that analyzed the SMEs' awareness and experience with EAM and intended to capture recent initiatives in this topic. This work reveals a need for a reference enterprise architecture (EA) that tailors utility enterprises demands towards EAM and derives implications for the development of such a reference EA.

Keywords: Enterprise Architecture Management, Reference Modelling, Utility Industry, Energy Market, Small and Medium-Sized Enterprises, Reference Enterprise Architecture, Business-IT-Alignment, Survey

1 Introduction

Enterprises are complex and highly integrated systems comprised of processes, organizations, information and technologies, with interrelationships and dependencies in order to reach common goals [RSB11]. They need to be aware of the relations among strategy, business processes, applications, information infrastructures and roles to rapidly react on changing demands in the market and within their organization. Enterprise Architecture Management (EAM) contributes to this purpose by providing methods and tools to establish a more holistic perspective on enterprises [Ahl12, Lan13] which includes systematically capturing and developing the different architectural layers of an enterprise (e.g. business, application and technology architecture).

Recent developments like market liberalization and the diversification of energy sources caused significant changes in the European utility industry [JKK11]. Numerous new market roles and business opportunities created by changes in regulations resulted in an

[1] University of Rostock, Chair of Business Information Systems, Albert-Einstein-Str. 22, 18059 Rostock, felix.timm@uni-rostock.de
[2] University of Rostock, Chair of Business Information Systems, Albert-Einstein-Str. 22, 18059 Rostock, matthias.wissotzki@uni-rostock.de
[3] University of Rostock, Chair of Business Information Systems, Albert-Einstein-Str. 22, 18059 Rostock, christina.koepp@uni-rostock.de
[4] University of Rostock, Chair of Business Information Systems, Albert-Einstein-Str. 22, 18059 Rostock, kurt.sandkuhl@uni-rostock.de

increased competition. Utility enterprises are forced to adapt their business models to the changing market situations which also requires adaptation in the enterprise architecture. In this context, EAM is expected to be an important tool for supporting change processes and developing competitive business capabilities [CWS13]. Especially small and medium-sized enterprises (SME) require support in EAM implementation, as they do not have a well-established tradition in this field [WTS15].

As a contribution to the above challenges, the objective of the ECLORA project is to develop a reference Enterprise Architecture for small and medium-sized enterprises in utility industry. This reference architecture can be applied in specific SMEs from the utility industry to help tailoring their actual enterprise architecture to their business demands depending on its characteristics like taken market roles. The focus of this paper is on utility enterprises' current awareness and knowledge of EAM concentrating on SMEs in Europe. Furthermore, the actual degree of EAM integration in these organizations is analyzed. For this purpose, an online survey was developed systematically questioning practitioners from the utility industry. The work is guided by three research questions (RQs):

RQ1: *To which extend are SME utility enterprises aware of EAM and its benefits?*

RQ2: *What is the current state of EAM initiatives in utility enterprises?*

RQ3: *Is there a demand for a reference Enterprise Architecture in Utility Industry?*

After clarifying the theoretical background regarding SMEs, EAM and the utility industry in section 2, the ECLORA project is introduced in section 3. Section 4 presents the survey design and evaluation of the responses, before in section 5 implications for ECLORA are derived. Section 6 finally concludes the contributions of this work and presents future actions to be taken for developing a reference EA.

2 Theoretical Background

Before presenting the survey's results, its theoretical foundation is clarified. The notion of SMEs is explained and how this work understands it. The field of research is EAM, whose current relevance for SMEs is elaborated. As the considered industry sector the SME utility industry is introduced and examined regarding EAM.

2.1 Small and Medium-Sized Enterprises (SME)

Depending on the context, enterprises can be classified by quantitative and qualitative criteria. According to [Cas02] more than 99% of European enterprises operate as an SME, which employ around 65 million people in total. Globally between 40% and 50%

of gross domestic product is accounted to SMEs. The European Commission defines enterprises up to 250 employees with a yearly business volume of up to 50 million Euro or a balance sheet total of up to 43 million Euro as SMEs [EU01]. This paper uses the less restrictive definition of medium-sized enterprises by the German institute for SME research. Here, enterprises are considered medium-sized with less than 500 employees [IfM01].

The entrepreneur and owner of a small enterprise is habitually working in his own company, therefore decision processes of SME are mostly highly centralized. They are often more focused on the operational planning and controlling than strategic aspects. The use of controlling instruments is usually limited. In addition, there is a big difference between the knowledge about the importance of such management models and the implementation of concrete measures [Kar11].

2.2 Enterprise Architecture Management and its Part in SMEs

Architecture is defined as a fundamental organization of a system embodied in its components, their relationships to each other, and to the environment, and the principle guiding the organizations design and evolution [Lan13]. EA is the formal declaration of the basic structures of an organization, its components and relations, as well as the processes used for development [WA12]. In this context, EAM provides a powerful approach for a systematic development of the organization in accordance with its strategic visions, yet its value depends on the organizational ability to perform EAM effectively [Ahl12].

The discussion and usage of EAM is focused on IT in practice, although there is knowledge about the importance of processes, strategies and organizational aspects as well [WA12]. EAM adopts various perspectives. As a *management philosophy*, it is a holistic way to understand, plan, develop control and adjust organizations architecture. As an *organizational function* it enables and improves existing strategic planning and implementation processes. As a *methodology and culture* it represents an open approach among the managers and proposes a set of management practices in order to reach a global optimum for the firm, free of egoism and opportunism [Ahl12].

Regarding EAM in SME there only has been few research so far [WA12]. In order to support the transition process of growing SMEs, [Jac11] developed the SME EA Growth Model (SMEAG) by dint of case study research. The model enhances existing growth stage models by combining EA principles for change management, EA frameworks and operational models for business execution. By mapping several areas of concern (e.g. organizational structure) to stages of growth, the work of [Jac11] reveals that EA facilitates growth triggered transition for SME. The need for standardizations and integration of processes in the several growth stages becomes more transparent. That motivates this work to contribute in research for EAM in SMEs.

2.3 The Utility Industry and Current State of EAM

Over the last two decades the European energy market has faced fundamental structural changes [JKK11]. Next to climate policies other regulations permanently changed the market utility enterprises operate in. In Germany within the EnWG (Energy Industry Act)[5] law market functions were legally separated from each other through unbundling, e.g. separating energy production, energy trade or energy transmission and distribution. Next this, also technical improvements increased competition, which forced utility enterprises to improve their efficiency and effectiveness [AC07, GSA12]. Further the German energy market contains numerous market roles with different responsibilities and functionalities. The Germany Federal Association of the Energy and Water Industry categorizes nine market roles such as energy retailer, balance grid coordinator or metering service provider [BDE08]. Since several roles can be taken by one utility enterprise, the organizational structure of one enterprise is varying from another.

From IS perspective today's utility enterprises have more complex requirements towards its information systems. In [GSA12] the authors identified more than 80 different information sources that have to be used in order to develop an appropriate information system for the utility industry. The authors of this work determine EAM as an approach facilitating business and IT compliance on the one, and optimization of organizational structures on the other side. The emerging objectives to align business and IT, to overcome IT complexity, and to reduce costs for sustain competitiveness can be reached by implementing EAM [Ahl12].

There has been activity in IS research, which addressed this issue. For instance, [GSA12] identified 11 reference models for information systems development in utility industry and proposed a catalogue for reference models in order to agree on a common terminology [GA10]. As a summary the authors derive a lack of current research regarding EAM initiatives in the utility industry [CWS13]. Although stated literature and projects address the role of information systems in this industry, a holistic approach like EAM cannot be identified. Especially SME utilities seem to demand EAM integration in order to cope with the mentioned developments. Thus, the ECLORA project aims at developing a reference EA for SME utilities that can be tailored to the enterprise at hand by configuring the reference EA during its application.

3 The Joint Project ECLORA

The contribution of this work was elaborated in the frame of the ECLORA (German acronym for *Development of a Cloud based Reference Architecture*) project. In this joint project the academic partner collaborates with an industrial partner that provides software solutions for SME utilities. The objective of ECLORA is to develop a reference EA for SME utility enterprises such as energy providers of a municipality. This intention

[5] cf. http://www.gesetze-im-internet.de/bundesrecht/enwg_2005/gesamt.pdf, accessed 2015-04-02

is facilitated by dint of reference modelling. Reference models are information models, developed for an abstract class of application and entitled to universality in this class. Thus, their purpose is to be reused by mechanisms of adjustment and extension according to a special application case. The reuse of a reference model is intended to increase both efficiency and effectivity of an enterprise's information systems and their change management [Bec02, FL04, Bro03].

ECLORA intends to develop a reference EA that uses the TOGAF framework as an EAM standard and thus will include all sub-architectures, namely business, information systems and technology architecture [TOG11].

From a methodical perspective the project utilizes the design science approach in order to develop the named artefact (i.e. the reference EA) by iteratively approximating it to a real world application and thereby enhancing it [Hev04]. For validation purposes technical action research is integrated in this iterative development process [WM12]. In Fig. 1 the instantiation of the applied regulative cycle in ECLORA is illustrated. As indicated, the survey presented in this work contributes both to the problem investigation and the solution design. Hence, the survey is intended to provide results conducing to an initial version of the reference EA, the focus of the work at hand is the analysis of the survey regarding the current state of EAM in utility industry.

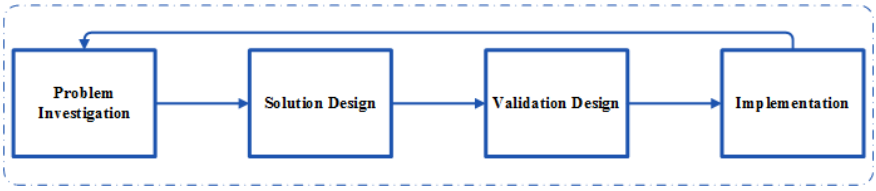


Fig. 1: Instantiating Design Science in ECLORA

4 Conducting a Survey in SME Utilities

In order to capture the current state of EAM in SME utilities a survey was conducted. The results are intended to answer the RQs of this work and reveal implications for the ECLORA project. Section 4.1 therefore explains the survey design before section 4.2 analyses its results and answers the RQs.

4.1 Survey Design

The scope of the survey was to collect information regarding EAM in SME utility enterprises. Therefore a questionnaire was designed and utilized by dint of an online survey. Surveys provide a simple approach to retrieve a high amount of information in a standardized way [Rob11]. In terms of design and conduction the steps carried out by [Rob11] guided the survey design. After revising the questionnaire a pre-test was

conducted with both partners of the joint project. The improved and condensed questionnaire was uploaded to the platform *SoSci Survey*[6] and its link was published via the project partner SIV.AG to customers from the SME utility industry. In the end 53 completely answered questionnaires have been submitted. With the export functionality of *SoSci Survey* (in CSV-format), the data was analyzed using SPSS[7] by IBM [Cam14].

Next to an introductory part providing basic information, the questionnaire contains six parts: *General Information (Part A)* focuses on the responding enterprise, its size, market role in the utility industry and the respondents' position within the enterprise. *Corporate Management and Organization (Part B)* focuses on knowledge and experiences in EAM as well as EAM's current integration and its general importance for the utility enterprise. *Strategy (Part C)* intends to capture the transparency of the enterprise's strategy, identifies core processes as well as Business Process Outsourcing (BPO) activities. *Operation (Part D)* enquires the enterprise's current IT landscape and its alignment to business and strategy. Further, the respondents' satisfaction of the current IT support is captured. *Monitoring (Part E)* addresses whether utilities implement processes for updating IT and business with operative and strategic changes and asks for utilized measuring methods. *Development and Forecast (Part F)* identifies future actions regarding EAM and asks for the expected value of implementing EAM.

4.2 Discussion of Results

This section presents the results after analyzing the 53 responded questionnaires. The discussion's structure is guided by the parts enumerated in section 4.1 and provides the foundation for deriving implications for ECLORA in section 5.

General Information (Part A): While SME are featured with 19 respondents, 15 employees from major enterprises answered the questionnaire. With 34% the majority of the respondents work in the management. The second big group of respondents (20.8%) came from the IT sector of their enterprise. Further respondents came from Marketing and Sales (13.2%), Accounting (7.5 %), Grid Operations and Services, Technics (both 5.7%), Procurement (3.8%), Corporate Development (1.9%) and Other (no exact statement, 5.7%).

The respondents were able to choose from nine different market roles in utility industry. These roles were derived from [BDE08] and adjusted to the target respondent group, i.e. end consumer or legislative institute were excluded. Since multiple answers were possible, 25 different combinations of market roles were identified in the pool of responding enterprises. Almost every second market role combination has less than 5% share of the total amount of combinations. This indicates a high degree of diversification in utility industry in terms of market role diversification. The diversification is illustrated by Fig. 2.

[6] cf. https://www.soscisurvey.de/
[7] cf. http://www-01.ibm.com/software/de/analytics/spss/

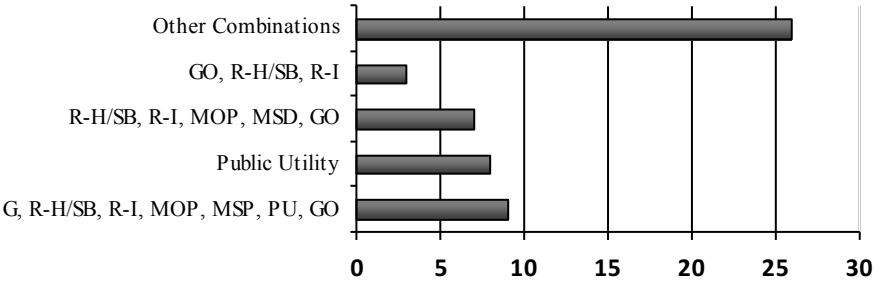


Fig. 2: Combination of Market Roles (G- Generator, GO - Grid Operator, MOP - Metering Point Operator, MSP - Metering Service Provider, PU - Public Utility, R-H/SB - Retailer for Households and Small Businesses, R-I - Retailer for Industry)

Corporate Management and Organization (Part B): Every respondent assessed his knowledge and experience with EAM as well as his opinion about EAM's importance for an enterprise. None of the respondents rated himself as an expert in EAM practice. Based on all respondents the big majority (61%) is inexperienced with EAM. Still, 72% of the IT respondents are at least beginners. 43% of Marketing and Sales as well as 39% of the management respondents consider themselves as beginners. Considering the amount of respondents, it can be derived that management and IT are the two most EAM aware and experienced departments. 50% of the management rated EAM as important or very important for enterprise success, while 82% of the IT department answered the same. In contrast to the missing experience of EAM, one third of every identified department rated EAM as important. This result is remarkable when contrasting the level of experience to the assessment of EAMs importance, which is visualized in Fig. 3. The variable importance is scaled since the amount of respondents differs per level of experience. With increasing experience of EAM, respondents assess EAM as more important.

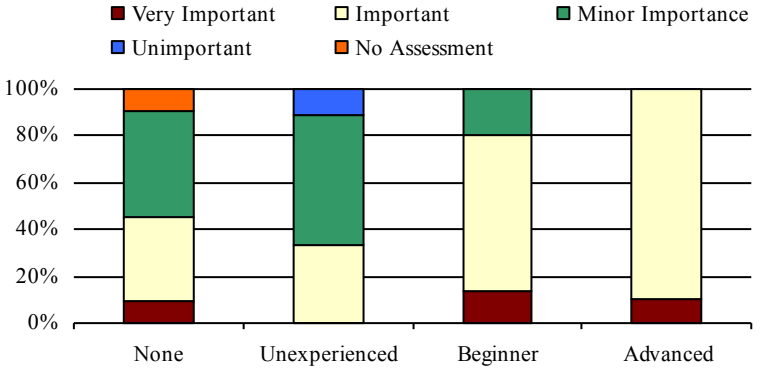


Fig. 3: Importance of EAM by Level of Experience

Only in 41% of the cases the respondents felt able to determine whether EAM has complete management support in their enterprise. While 28% declared they do have support, 13% said they do not. The following three reasons arose for the latter one: *EA Frameworks like TOGAF are too complex*, *EAM is too cost-intensive* and *EAM is too time-consuming*. This indicates, that research may lack in EAM solutions or frameworks tailored for SME. The statistic of EA Framework usage confirms this statement. 77% of the respondent did not make a declaration regarding used EA Frameworks. While COBIT or TOGAF only were mentioned by 5.7%, ITIL is used by 15.1%. Furthermore, this part of the questionnaire indicates what the respondents consider as the most promising benefits when implementing EAM in their enterprises. Especially overcoming complexity (45.3%) and tool support for business process management (41.5%) were mentioned. 35.8% stated that decision support for strategic projects and investments are promising benefits by EAM activities. Likewise advantages like quality management and the review of IT systems and applications were named by 32.1% resp. 30.2%. Further, the respondents had to state their feelings about the following three statements: *(1) Acquisitions and fusions force the harmonization of heterogeneous IT landscapes, (2) Movements like Cloud Computing, Offshoring or Outsourcing lead to a distributed IT landscape. (3) Regulations and laws cause an increasing IT complexity.* The agreements for these statements were measured using a Likert scale [Nor10]. The result revealed that the majority agreed to all three statements by choosing at least "agree". In consequence, all these three aspects seem to characterize current developments in utility industry and could be addressed by dint of implementing EAM in SMEs.

Strategy (Part C): In part C the respondents answered questions regarding documentation strategies in their enterprises. 70% assessed their corporate strategy as documented, of which only 49% stated it as available for every employee. 17% named the corporate strategy as undocumented. Further, Business Goals [San14] were documented in 81% of the cases, of which only 58% had access to it. In 7.5% of the cases business goals were not documented at all. The third aspect was the documentation strategy of core processes. They are key activities that must be performed to ensure the enterprises continued competitiveness since it adds primary value to a certain outcome. From 84.9% of documented core processes 53.3% were accessible by the employees. This leads to the statement, that although most enterprises document their strategies, goals and processes, more efforts have to be made in order to provide a transparent strategy, the basis for successful EAM [Ahl12].

The questionnaire further investigated the aspect of core processes in the utility industry. Although 91% of the respondents said that the core processes are supported by information systems, only 22% stated to be delivered sufficiently by enough information and resources. It seems that there is a need for optimizing IT support of the core processes. This statement is supported when contrasting the degree of IT support and the provision of information and resources.

The last aspect was the development of Business Process Outsourcing (BPO). Each respondent was able to choose multiple from given business functions and their sub-

functions. Especially Energy Data Management (28%) and Billing Processes (20%) seem to be frequently outsourced processes in utility industry. Further, Grid Operations (16%) and end-consumer processes (11%) seem to be often outsourced. In general BPO can be stated as an important topic and need to be considered during the development of a reference EA in utility industry.

Operation (Part D): This part focused on Business and IT alignment in the respondents' enterprise. For the analysis of the responses also the results from part C were brought in here. The respondents had to assess to what extend IT and functional departments of their enterprises are collaborating. The in part C identified need for improvement of IT support was not verified at this point. 47.2% of the respondents stated a complete collaboration between IT and functional departments. 49.1% said the collaboration partially existed. Contrasting these results with the questions from part C did not provide an explanation why information and resources provision is lacking. A conclusion of these results could be that even with high collaboration of IT and business, the IT support and resource provision are not automatically satisfying. Thus, the practice of IT and business collaboration might not be optimal and need to be improved.

The central question was to name applications and software solutions related to the given business functions. This contributed especially for elaborating the interrelations among business and information architecture of the TOGAF Framework in ECLORA, which is out of the scope of this work. Generally it can be said that primarily commercial software suites and solutions were used as IT support. SAP, Neutrasoft and Schleupen were identified as the most utilized IT providers.

Monitoring (Part E): In order to capture changes in the enterprise architecture, monitoring is a vital part of EAM. Changes in business processes or responsibilities may have direct impact on operational business and thus need to be identified in order to adjust the EA [Ahl12]. Nevertheless, in 47% of the surveyed enterprises neither a single person nor a department is responsible for this interface activity between IT and business. Only 25% of the enterprises have a responsible role and 11% a department assigned for this.

Most of the enterprises update their corporate strategy, IT strategy and core processes in irregular time intervals. While 40% update their corporate strategy on demand, about 60% analyze their IT strategy and core processes if required. Only 28% of them update their corporate strategy annually. Almost every (96%) respondent stated that no particular measuring methods are used for monitoring. This allows the presumption that monitoring in current utility industry is very reactive instead of being proactive, which would be demanded by EAM.

Development and Forecast (Part F): The last part evaluated in which business functions the respondents ask for simpler IT support. More than 25% mentioned to be satisfied with the current situation. Still, 55% marked several business functions. Especially billing systems, customer relationship management systems as well as document management systems seem to be too complex. Further, it had to be stated which business

functions would get more efficient with better IT and business alignment. This intends to make expectations regarding EAM transparent. For each of the selectable business function multiple sub-functions could be chosen. It can be seen that especially billing processes (23%), end-consumer processes (20%) as well as controlling processes (15%) and energy data management (14%) seem to be expected to benefit most from EAM.

Next to the estimated impacts of EAM, respondents were asked to state which business functions tend to be outsourced in long-term. 86.8% assess current BPO initiatives as sufficient. There is no business area which tends to be generally outsourced in the future of utility industry additionally to the stated ones in Part C.

The last question intended to capture the expectations of utility industry regarding EAM. Again, multiple selections were possible. The following list summarizes the most marked expectations: *more effective utilization of resources* (53%), *detection of demand for action and optimization* (49%), *enhancement of business IT alignment* (43%), *analysis-tool to understand relationships and dependencies in the enterprise* (36%), *reduction of IT costs* (32%).

These results go along with the findings in Part B, where more than 45% said that overcoming IT complexity is enabled by implementing EAM in their enterprises. The results of this part indicate that especially in terms of billing and end-consumer processes EAM will facilitate an optimized resource provision, an enhanced Business-IT-Alignment and cost reduction. Likewise this could apply for Energy Data Management and Controlling Processes.

5 Implications for ECLORA

The analysis of the survey revealed seven major findings that are listed below. For each finding a consequence for the project is derived. The part of the questionnaire, where the finding occurred is indicated in brackets.

I. High Diversification of Market Roles in Utility Industry (Part A): In total, 25 different combinations of market roles were identified. This implies that a utility enterprise's EA partially depends on the market roles it takes. Thus, when identified a market role of the enterprise, configuration rules can be applied in order to tailor the reference architecture to the enterprise.

II. The assessment of EAM's importance for success increases with growing awareness and knowledge of the EAM discipline (Part B): This reveals that utility companies have to be made aware of EAM in order to convey its importance to the management level. Thus, when conducting workshops in ECLORA, knowledge of EAM has to be assessed and contingently imparted to the enterprise.

III. SME Utility Enterprises find EA Frameworks too complex and expensive for their business (Part B): Although there are numerous EA Frameworks available, SME utilities

do not feel supported by them since they address bigger business structure. Thus, there is a lack of EA frameworks that are tailored or can be scaled down to small and medium-sized enterprises. The reference architecture developed in ECLORA tries to address this issue and adjust the TOGAF Framework for SME utilities.

IV. Demand for a Reference EA was validated by the survey (Part B and D): The survey identified three factors that let utilities' EA grow complex. Next to fusions and outsourcing strategies, especially rules and regulations require an advanced flexibility from utilities (68% of the respondents). Further, almost the half of the EAM experienced respondents characterized EAM as a measure to overcome complexity. Together with finding (III) this underlines the demand of SME utilities for EAM.

V. Potential Optimization of IT Support and Communication between IT and other departments (Part C): Although in every case the identified core processes were supported by IT, the majority of respondents neither felt sufficiently delivered with information and resources nor are satisfied with the IT support. This reveals insufficient Business-IT-Alignment. This is addressed by ECLORA by mapping the business architecture with the other layers of the TOGAF Framework.

VI. Business Process Outsourcing is a relevant topic in Utility Industry (Part C): Especially in Energy Data Management and Billing the responding enterprises utilized outsourcing strategies. In order to provide a profound reference EA, mechanisms for configuration need to be developed, that address outsourcing strategies. Further, strategic changes regarding outsourcing have to be dealt with in a flexible manner applying the reference EA.

VII. Monitoring Processes need to be implemented in Utilities (Part E): In almost half of the cases Monitoring was conducted by not even a single person. Further, updating corporate and IT strategy as well as documentation of core processes were revealed as immature. The same applies for measuring methods. Thus, the reference EA also needs to address the implementation of monitoring in order to move utilities from a reactive towards a proactive manner.

6 Conclusion and Outlook

The aim of EAM is to master the complexity of IT and to align it to the enterprise's objectives, its business and also other aspects like laws or regulations [Ah12]. Especially utility industry is expected to be a beneficiary of the integration of EAM since laws such as market liberalization require utilities to act competitive. In the frame of the ECLORA project, a reference EA is developed, which applies reference modelling in order to provide a universal solution for EAM integration in utility industry [Ard13].

Therefore this work analyzed EAM's current state in the utility industry focusing on SMEs. A questionnaire was designed in order to capture utility enterprise's current

awareness of EAM and its benefits (RQ1), present EAM initiatives and resulting experiences (RQ2) and to reason whether there is a demand for a reference Enterprise Architecture in SME utility enterprises (RQ3).

The survey reached 53 respondents, which came from the utilities' management or IT department mainly. Facing RQ1 only respondents from IT can be considered as experienced with EAM. Apart from this, the majority ranked EAM as important for corporate success and named present issues like IT complexity as benefits when integrating EAM. Regarding RQ2 only few and no holistic EAM initiatives were was identified. Only a minority of respondents felt sufficiently provided with resources and information and are satisfied with IT support. Further a lack of monitoring actions was captured since in the majority of cases no responsibilities are assigned for monitoring and updating the EA. Although these findings reveal the importance of EAM, common EA frameworks like TOGAF were stated as too complex and cost-intensive for SMEs to implement. Laws and regulations were assessed as IT complexity drivers as well as numerous BPO initiatives cause an increasing heterogeneity of IT. Further, the high diversification of market roles in utility industry accompanies variations of the utilities' corporate strategy and thus their EA. In consequence, a demand for a reference EA in SME utilities was identified (cf. RQ3).

The demanded reference EA needs to be configurable to its application context. Hence, ECLORA will develop the reference EA applying the approach for configurable reference modelling [Bec02]. The implications stated in section 5 will guide the next steps in the solution design phase (cf. section 3): The development of an initial reference EA, based on the results presented, and an additional literature review. This reference will then be validated and redesigned applying it to numerous utility enterprises by conducting multi-day workshops.

References

[AC07] H.-J. Appelrath and P. Chamoni, "Veränderungen in der Energiewirtschaft — Herausforderungen für die IT," Wirtsch. Inform, vol. 49, no. 5, pp. 329–330, 2007.

[Ahl12] F. Ahlemann, E. Stettiner, M. Messerschmidt, and C. Legner, Strategic enterprise architecture management: Challenges, best practices, and future developments. Berlin, New York: Springer, 2012.

[Ard13] P. Ardalani, C. Houy, P. Fettke, and P. and Loos, "Towards A Minimal Cost Of Change Approach For Inductive Reference Model Development," in ECIS, 2013.

[BDE08] BDEW, "Leitfaden - Marktzugang für neue Teilnehmer", 2008.

[Bec02] J. Becker, P. Delfmann, R. Knackstedt, and D. Kuropka, "Konfigurative Referenzmodellierung", in Wissensmanagement mit Referenzmodellen: Konzepte für die Anwendungssystem- und Organisationsgestaltung, Physica-Verl, 2002, pp. 25–144.

[Bro03] J. Vom Brocke, Referenzmodellierung: Gestaltung und Verteilung von Konstruktionsprozessen. Berlin: Logos-Verl, 2003.

[Cam14] P. Cammin, “Entwicklung eines Rahmenwerkes zur Analyse von Unternehmens-architekturen in der Versorgerindustrie”, Master's Thesis, University of Rostock, 2014.

[Cas02] C. Cassell, S. Nadin, M. Gray, and C. Clegg, “Exploring human resource management practices in small and medium sized enterprises,” Personnel Review, vol. 31, no. 6, pp. 671–692, 2002.

[CWS13] C. Czarnecki, A. Winkelmann, and M. Spiliopoulou, “Reference Process Flows for Telecommunication Companies,” Bus Inf Syst Eng, vol. 5, no. 2, pp. 83–96, 2013.

[EU01] The European Commission, What is SME?: User Guide and Model Declaration,2015.

[FL04] P. Fettke and P. Loos, “Referenzmodellierungsforschung,” Wirtschaftsinf, vol. 46, no. 5, pp. 331–340, 2004.

[GA10] González J. M. and H.-J. Appelrath, “Energie-RMK - Ein Referenzmodellkatalog für die Energiewirtschaft,” in Modellierung 2010, G. Engels, D. Karagiannis, and H. C. Mayr, Eds, Karlsruhe: GI, 2010, pp. 319–334.

[GSA12] González Vázquez, José Manuel, J. Sauer, and H.-J. Appelrath, “Methods to Manage Information Sources for Software Product Managers in the Energy Market,” Bus Inf Syst Eng, vol. 4, no. 1, pp. 3–14, 2012.

[Hev04] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design science in Information Systems research,” MIS QUARTERLY, vol. 28, no. 1, pp. 75–105, 2004.

[IfM01] Institut für Mittelstandsforschung Bonn. Available: http://www.ifm-bonn.org/ (2015, Apr. 01).

[Jac11] D. Jacobs, P. Kotzé, A. van der Merwe, and A. Gerber, “Enterprise Architecture for Small and Medium Enterprise Growth” , EEWC 2011, LNBIP 79, Springer, 2011, pp. 61–75.

[JKK11] U. C. C. Jagstaidt, J. Kossahl, and L. M. Kolbe, “Smart Metering Information Management,” Bus Inf Syst Eng, vol. 3, no. 5, pp. 323–326, 2011.

[Kar11] D. Kardel, “IT-Sicherheitsmanagement in KMU,” HMD, vol. 48, no. 5, pp. 44–51, 2011.

[Lan13] M. Lankhorst, Enterprise architecture at work: Modelling, communication and analysis, 3rd ed. Heidelberg, New York: Springer, 2013.

[Nor10] G. Norman, “Likert scales, levels of measurement and the "laws" of statistics,” (eng), Adv Health Sci Educ Theory Pract, vol. 15, no. 5, pp. 625–632, 2010.

[Rob11] C. Robson, Real world research: A resource for users of social research methods in applied settings, 3rd ed. Chichester, West Sussex: Wiley, 2011.

[RSB11] M. Razavi, F. Shams Aliee, and K. Badie, “An AHP-based approach toward enterprise architecture analysis based on enterprise architecture quality attributes,” Knowl Inf Syst, vol. 28, no. 2, pp. 449–472, 2011.

[San14] K. Sandkuhl, J. Stirna, A. Persson, and M. Wißotzki, Enterprise modeling: Tackling business challenges with the 4EM method, 2014.

[TOG11] The Open Group, TOGAF Version 9.1, 1st ed. Zaltbommel: Van Haren Publishing, 2011.

[WA12] M. Wißotzki and A. Sonnenberger, “Enterprise Architecture Management - State of Research Analysis & A Comparison of Selected Approaches,” in Short Paper Proceedings of the 5th IFIP WG 8.1: CEUR-WS.org, 2012.

[WM12] R. Wieringa and A. Moralı, “Technical Action Research as a Validation Method in Information Systems Design Science”, DESRIST 2012, LNCS 7286, Springer, 2012, pp. 220–238.

[WTS15] M. Wißotzki, F. Timm, and A. Sonnenberger, “A Survey on Enterprise Architecture Management in Small and Medium Enterprises,” in ICEIS, Barcelona, Spain, 2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

i*Gov a Feeling – Herausforderungen und Technologien zur Realisierung eines Studienportal für das interaktive Heute

Manuel Breu[1], Klaus Berndl[2], Thomas Heimann[3], Lars Santesson[4]

Abstract: Der Weg vom Studieninteressenten zum Studenten ist in Deutschland geprägt von redundanten Datenerfassungen, vielfältigen Anlaufstationen und für den Bürger komplex anmutenden Prozessen. Dieses Beispiel ist symptomatisch für viele digitale Leistungsangebote der öffentlichen Hand. Eine niedrige Akzeptanzrate des E-Governments in der Bevölkerung ist damit kaum verwunderlich. Das Rahmenwerk i*Gov liefert Konstruktionsprinzipien, um digitale Leistungsangebote so zu gestalten, wie sie der Bürger erwartet. Dieses Positionspapier skizziert ein interaktives, nutzerorientiertes Studienportal, das den Konstruktionsprinzipien von i*Gov folgt. Es zeigt Schlüsselelemente- und Technologien einer Lösungsarchitektur.

Keywords: eGovernment, i*Gov, Studienportal

1 Der steinige Weg vom Studieninteressenten zum Studenten

Auf ihrem Weg zu einem Studium, sehen sich Studieninteressenten in Deutschland mit verschiedensten Informationsquellen, Medienbrüchen und komplexen Prozessen konfrontiert. Im Internet existieren beispielsweise zahlreiche, heterogen gestaltete Informationsplattformen unterschiedlicher Anbieter, welche Informationen zu Studium und Studiengängen bereitstellen. Eine mögliche Anlaufstation zu Studieninformationen sind auch die Studienberatungen der Hochschulen. Diese sind oft nur telefonisch erreichbar. Bewerben sich Studieninteressanten an Universitäten, so müssen sie ihre Personendaten je Hochschule in unterschiedlich gestaltete Bewerbungsformulare eingetragen. Oft ist es anschließend notwendig, diese Formulare auszudrucken, zu unterschreiben und auf postalischem Weg an die Hochschule zu senden. Dieser Vorgang wiederholt sich bei der Immatrikulation. Eine persönliche Immatrikulation an der Hochschule ist vielerorts üblich. Möchten sich Studenten schließlich nach erfolgreicher Immatrikulation etwa für einen Wohnheimplatz bewerben, müssen Informationen über Wohnheime ebenfalls über verschiedene Informationskanäle und -plattformen beschafft werden. Eine Bewerbung geschieht hier oft über eigene Formulare auf den Internetseiten der Wohnheimbetreiber, per Email oder schriftlich per Post.

Diese Punkte sind nur Beispiele. Sie sind aber symptomatisch für viele Kundenprozesse der öffentlichen Hand in Deutschland: Redundante Datenerfassungen, Medienbrüche, und aus Bürgersicht komplexe Prozesse mit vielen Anlaufstellen sind allgegenwärtig.

[1] Capgemini Deutschland GmbH, Potsdamer Platz 5, 10785 Berlin, manuel.breu@capgemini.com
[2] Capgemini Deutschland GmbH, Potsdamer Platz 5, 10785 Berlin, klaus.berndl@capgemini.com
[3] Capgemini Deutschland GmbH, Potsdamer Platz 5, 10785 Berlin, thomas.heimann@capgemini.com
[4] Capgemini Deutschland GmbH, Potsdamer Platz 5, 10785 Berlin, lars.santesson@capgemini.com

Vergleicht man dies mit den digitalen Angeboten der Privatwirtschaft, scheint es, als habe der Staat unser digitales, kundenorientiertes Zeitalter schlicht verschlafen. Es stellt sich die Frage: Ist das wirklich so?

2 *i*Gov* weist den Weg vom elektronischen hin zum interaktiven Government

Wie das E-Government-Gesetz zeigt, beschäftigt sich die öffentliche Hand durchaus mit Digitalisierung. Dieses Gesetz verfolgt die Ziele, die elektronische Kommunikation mit der Verwaltung zu erleichtern und es Bund, Ländern und Kommunen zu ermöglichen, einfachere, nutzerfreundlichere und effizientere elektronische Verwaltungsdienste bereitzustellen [Bu14]. Ein Treiber dieser in Gesetzestext gegossenen Transformationsmaßnahmen sind Budgetsanierungen aufgrund von Haushaltskonsolidierung und Schuldenabbau, die mit der Alterung der Gesellschaft kollidieren. Darüber hinaus gehen viele Arbeitskräfte im öffentlichen Dienst in näherer Zukunft in Ruhestand. Der Anteil der über 50 jährigen liegt in Deutschland im Durchschnitt bei ca. 25% [Ro09]. Weil diese auf Grund von Haushaltsbeschränkungen nicht durch Neueinstellungen ersetzt werden, wird die IT der öffentlichen Verwaltung geradezu gezwungen, stark zu automatisieren [Re14]. Die britische Verwaltung verfolgt hier eine „*Digital by Default*"-Strategie und somit einen Push-Ansatz: Jegliche Interaktion geschieht elektronisch und nur in Ausnahmefällen durch persönlichen Kontakt [Of12]. Dies ist in Deutschland nicht gewollt. Nach Einschätzung der Politik birgt dies die Gefahr einer Verschlechterung auf der Leistungsseite, v.a. dort wo eine persönliche Kommunikation sinnvoll ist [BS13]. In Deutschland sollen digitale Leistungsangebote geschaffen werden, die der Bürger freiwillig nutzt [Re14]. Für den Erfolg einer solchen Pull-Strategie sind die Akzeptanzraten der digitalen Angebote jedoch zu gering: Nur etwa 34% aller deutschen Onliner nutzen E-Government-Dienste [Ro13]. Wie Leistungsangebote digitalisiert werden, folgt vor allem aus den Abläufen der Verwaltungen, nicht aber aus Sicht der Kundenprozesse. Ein Grund hierfür kann sein, dass verwaltungsübergreifende Dienste deutlich komplexer zu realisieren sind. Die Ergebnisse entsprechen daher oft nur bedingt den Erwartungen der Bürger [Re13]. Dies führt zu einer niedrigen Akzeptanz und kollidiert mit der Pull-Strategie Deutschlands. Noch expliziter als Ansätze wie Stein-Hardenberg 2.0 [Su11], und Verwaltung 4.0 [KH13] beschreibt Capgeminis Rahmenwerk i*Gov [Ro13] deshalb den Weg *vom e-Government zum i-Government*, d.h. von der behördlichen Elektronifizierung hin zur Interaktion. *i*Gov* kann als Beschreibungsrahmen verstanden werden, in welchen man Anwendungen oder Prozesse einordnen kann. So erleichtert es *i*Gov* diese Artefakte zu strukturieren und z.B. fehlende Elemente zu identifizieren. Dies hilft dabei, ein Vorgehen für die Schaffung einer integrierten Verwaltungsfabrik zu identifizieren. Sie kann dann als Basis für die Realisierung eines *intuitiven*, *individualisierten* und schließlich *interaktiven* Governments genutzt werden (vgl. Abbildung 1).

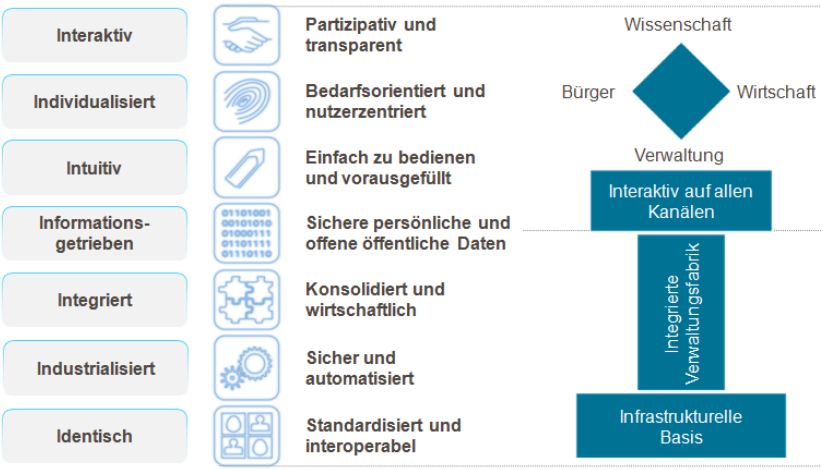


Abbildung 1 Struktur des i*Gov von Capgemini [Ro13]

3 Grobskizze eines nutzerorientierten Studienportals

In unserer digitalen Welt sind Portale zentrale Elemente der IT-Unternehmensarchitektur. Kunden können mittels Portalen jederzeit und an jedem Ort erreicht werden. Ein Zugriff ist mittels eines Webbrowsers möglich, welcher auf allen gängigen Plattformen (Desktop, Tablet, Smartphone,…) verfügbar ist. Darüber hinaus ermöglichen es Portale, Leistungsangebote nutzerorientiert zu integrieren – auch über Organisationsgrenzen hinweg. Als eine den Leistungsangeboten übergelagerte Klammer, können sie dabei insbesondere die Gestaltungsprinzipien interaktiv, individualisiert und intuitiv realisieren. Im Anwendungsfall eines Studienportals sollten Studenten so über ihre gesamte Lebenslage hinweg begleitet werden. Dies ist möglich, wenn Leistungsangebote um folgende Funktionen herum integriert werden:

Informieren: Allgemeine Informationen zu Themen rund um das Studium helfen Studieninteressenten, sich schnell in ihrer neuen Lebenslage zu Recht zu finden. Hierfür sind neben Informationen zu Abschlüssen, Hochschultypen usw. auch Anknüpfungspunkte zur Wohnungssituationen an Studienorten, Freizeitangebote, usw. sinnvoll. Informationen werden umfassend und strukturiert dargestellt, wie es der Benutzer intuitiv erwartet.

Suchen: Eine intelligente Suche nach Studiengängen und Hochschulen unterstützt bei der Wahl des richtigen Studiengangs. Ein Interessentest kann passende Studiengänge vorschlagen. Eine Volltextsuche mit individuell gewichtbaren Filtern (z.B. Studienort,…) hilft beim Stöbern nach Studiengängen. Personalisierte Vorschläge erleichtern die Suche. Bewertungen, Diskussionsforen und soziale Medien liefern

weiterführende Informationen.

Bewerben und Einschreiben: Bewerbung und Einschreibung für einen Studiengang sind elektronisch und medienbruchfrei möglich. Authentifiziert durch den neuen Personalausweises (nPA) [Bu15] können Benutzer zentral hinterlegte Personendaten in Bewerbungsformulare übertragen. Bei nicht zulassungsbeschränkten Studiengängen kann eine Einschreibung ad hoc geschehen.

Studieren: Den Studenten werden zentral Informationen zu Bafög, Stipendien, Auslandssemester, Praktikumsstellen, usw. bereitgestellt. Die Dateneingabe – zum Beispiel bei einer Bafög-Beantragung - wird analog zur Immatrikulation erleichtert. Auch kann das Portal eine zentrale Plattform für ein Engagement in Hochschulgruppen bieten.

Alle Funktionen sollen transparent auffindbar sein. Die Gestaltung des Portals soll in zeitgemäßem „Look-And-Feel" erfolgen. Eine ansprechende Darstellung im Webbrowser und Mobile-Endgeräten muss gewährleistet werden. Abbildung 2 zeigt, wie ein solches Studienportal aussehen könnte.

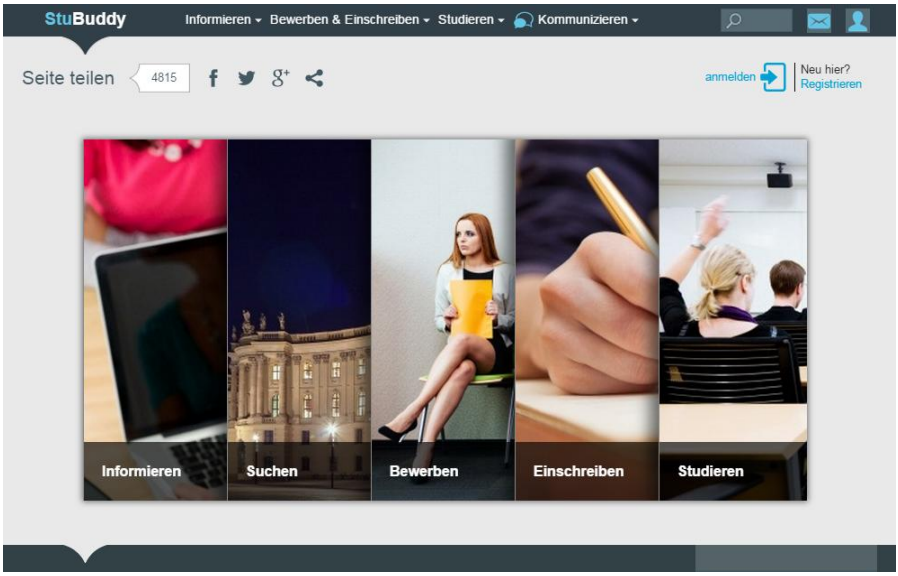


Abbildung 2 Vision eines Studienportals

4 Portaltechnologien ermöglichen ein i*Gov Studienportal

Leistungsangebote sollen um die Kernfunktionen Informieren, Suchen, Bewerben, Einschreiben und Studieren gebündelt und transparent auffindbar sein. Um solch ein

Studienportal zu schaffen, müssen vielschichte Herausforderungen bewältigt werden. Diese Herausforderungen werden insbesondere von den Gestaltungsprinzipien *informationsgetrieben*, *integriert*, *industrialisiert* und *identisch* des i*Gov (vgl. Abbildung 1) adressiert. Es existiert bereits eine Vielzahl verschiedener Anwendungssysteme unterschiedlicher Träger, welche schon heute z.T. große Datenmengen verwalten. Oft wurden hohe Summen für Insellösungen investiert. Die Anreize für die Betreiber, diese aufzugeben sind deshalb gering. Somit ist es kaum mach- und durchsetzbar, diese durch ein neues Studienportal zu ersetzen. Auch ist eine Integration bestehender Lösungen durch einen Open-API-Ansatz [Wi15] aufgrund der großen Anzahl beteiligter Instanzen – und des damit verbundenen immensen Organisations-, Kommunikations- und Abstimmungsaufwands - kaum umsetzbar. In Deutschland existieren derzeit über vierhundert staatlich anerkannte Hochschulen. Mit Portaltechnologien könnte dennoch, die vom i*Gov vorgeschlagene integrierte Verwaltungsfabrik realisiert werden. Sie dient als Basis des Studienportals. Denkbar wären dabei verschiedene technische Integrationsansätze. Folgende beispielhafte Technologien zeigen jedoch, dass auch hier Herausforderungen überwunden werden müssen.

iFrame-Integration [GB06]: Vorhandene Hochschulsysteme würden über iFrames in das Studienportal integriert. Dies ist eine der einfachsten technischen Integrationsmöglichkeiten. Hiermit ist es jedoch kaum möglich, ein gemeinsames „Look and Feel“ aller Systeme zu erreichen.

Reverse-Proxy Entwurfsmuster [Li15]**:** Ein Reverse-Proxy wirkt als Stellvertreter zwischen Client und einem Drittsystem – etwa einer Hochschulseite. Er nimmt http-Requests der Clients an, und leitet sie an das Drittsystem weiter. Der Response der Zielanwendung wird dann im http-Proxy so gefiltert, dass darin enthaltene Verweise wieder das Studienportal referenzieren. So wird der Kommunikationsweg über den Proxy auch bei zukünftigen http-Requests, die auf den initialen Request aufbauen, nicht verlassen. Abbildung 3 zeigt dies vereinfacht.

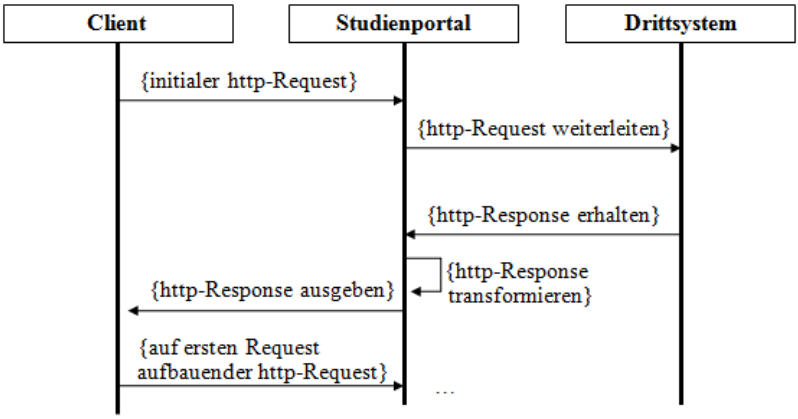


Abbildung 3 Reverse-Proxy: Grobe Funktionsweise

Gegenüber einer Einbindung von Drittsystemen mit iFrame-Integrationsmechanismen, bietet dieses Vorgehen einen wichtigen Vorteil: Neben den in der http-Ressource enthaltenen Verweise können auch die Inhalte an zentraler Stelle transformiert werden, etwa um sie dem „Look-And-Feel“ des Studienportals anzupassen. Hierfür müssten jedoch im Extremfall mindestens eine oder mehrere Transformationsregeln je Hochschulseite erstellt und verwaltet werden.

Web Services for Remote Portlets (WSRP) [Wi14]**:** Mit WSRP können Service und Benutzeroberfläche gemeinsam als grobgranulare Bausteine in einem übergelagerten Meta-Portal integriert werden. „Look-And-Feel“-Transformationen werden von diesem Standard grundsätzlich unterstützt. WSRP ist jedoch nur anwendbar, wenn die integrierten Drittsysteme selbst schon Portaltechnologien verwenden und diesen Standard unterstützen. Bei vielen bereits vorhandenen Systemen ist dies jedoch nicht der Fall.

Solange die organisatorischen Rahmenbedingungen für ein solches Studienportal nicht geklärt sind, hat eine solche Auflistung möglicher Technologien nur den Charakter eines Fingerzeigs. Wichtiger ist es zunächst, Verantwortlichkeiten zu klären. Die Schaffung eines übergreifenden Architekturmanagements für Hochschulsysteme ist nicht zuletzt schwierig, weil die Zuständigkeiten für Bildung in Deutschland föderativ geteilt sind. Zentrale Gremien wie die Hochschulrektorenkonferenz können hier beispielsweise ein Ansprechpartner und Treiber von Veränderungen sein.

5 Zusammenfassung und Ausblick

Wie das eingängige Beispiel zeigt, reichen die bestehenden Ansätze nicht, um eine freiwillige Nutzung des digitalen Kanals zu erreichen. Das vorgeschlagene Studienportal

ist nur ein Muster, das mit dem Rahmenwerk i*Gov auf andere Prozesse der öffentlichen Verwaltung übertragen werden kann. Der aufgezeigte Portalweg könnte hier eine Lösung sein, bestehende Systeme minimalinvasiv zu verändern und gleichzeitig mehr digitale Prozesse nutzerfreundlich anbieten zu können. Portaltechnologien können wichtige Enabler sein, den Portalweg – trotz schwieriger Rahmenbedingungen - gangbar zu gestalten.

Es wäre sehr wünschenswert, dass die öffentliche Verwaltung entsprechende Projekte in naher Zukunft realisiert. Sie können als Leuchtturm dienen. Einerseits weisen sie die Richtung für die Weiterentwicklung des E-Government hin zu nutzerorientierten, interaktiven Lösungen. Andererseits schaffen sie Sichtbarkeit beim Bürger. Sie zeigen ihm, dass der Staat das digitale Zeitalter nicht verschlafen hat; dass die Interaktion mit den staatlichen Organen nicht komplex, zeitraubend und unübersichtlich sein muss. Eine stärkere Fokussierung der Nutzer fördert die Motivation der Bürger, digitale Leistungsangebote auch tatsächlich stärker zu nutzen und stellt einen Erfolg der deutschen e-Government-Strategie sicher – vorausgesetzt sie werden auch richtig vermarktet. Dies sichert nicht zuletzt den Industriestandort Deutschland.

6 Literaturverzeichnis

[BS13] Breu, M.; Santesson, L.: EAM öffnet Sozialversicherungen den Weg in die digitale Zukunft. Berliner Behördenspiegel 01/2013, 2013.

[Bu14] Bundesministerium des Innern. E-Government-Gesetz, 2014, http://www.bmi.bund.de/DE/Themen/IT-Netzpolitik/E-Government/E-Government-Gesetz/e-government-gesetz_node.html, Zugriff am 10.04.2015

[Bu15] Bundesministerium des Innern. Personalausweisportal - Fragen und Antworten. http://www.personalausweisportal.de/DE/Home/home_node.html, 2015, Zugriff am 17.03.2015.

[GB06] Gornitsky, Richard; Boezeman, John: Options for rapid integration of Web applications into WebSphere Portal, http://www.ibm.com/developerworks/websphere/library/techarticles/0607_boezeman/0607_boezeman.html, 2006, Zugriff am 15.04.2015.

[KH13] Kruse, W.; Hogrebe, F.: Verwaltung 4.0- Erste empirische Befunde, http://subs.emis.de/LNI/Proceedings/Proceedings229/29.pdf, 2015, Zugriff am 09.05.2015.

[Li15] Liferay, http://www.liferay.com/community/wiki/-/wiki/Main/Web+Proxy+Portlet, 2015, Zugriff am 15.04.2015.

[Of12] Office UK Cabinet: Government Digital Strategy. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/296336/Government_Digital_Stratetegy_-_November_2012.pdf, 2012, Zugriff am 17.03.2015, S. 2-4.

[Re13] Reinhardt, M. Nachgefragt: Sind E-Government-Services noch zu wenig

nutzerorientiert Innovative Verwaltung 09/2013, Gabler-Verlag, 2013.

[Re14] Reinhardt, M.: European Executive Panel - Drivers for Digital Change. http://mc.capgemini.de/wfb2014/Plenary_Drivers_of_change_MarcReinhardt.pdf, 2014, Zugriff am 17.03.2015, S.2-6.

[Ro09] Robert Bosch Stiftung: Demographieorientierte Personalpolitik in der öffentlichen Verwaltung, 2009, http://www.bosch-stiftung.de/content/language2/downloads/Demographieorientierte_Personalpolitik_fuer_Internet.pdf, 2009.

[Ro13] Roth, S. L.: Workshop ZUKUNFTSKONGRESS – Mit sieben i zum bürgerorientierten E-Government. http://www.de.capgemini.com/zukunftskongress, 2013, Zugriff am 17.03.2015.

[Su11] Schuppan, T.: „Stein-Hardenberg 2.0“? – Staatsmodernisierung, Territorialität und verwaltungswissenschaftliche Transformationsforschung, 2011, http://www.budrich-journals.de/index.php/dms/article/view/5687/4815, Zugriff am 08.06.2015

[Wi14] Wikipedia: Web Services for Remote Portlets, https://en.wikipedia.org/wiki/Web_Services_for_Remote_Portlets, 2014, Zugriff am 15.04.2015

[Wi15] Wikipedia: Open API, http://en.wikipedia.org/wiki/Open_API, 2015, Zugriff am 15.04.2015

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Prozessorientierte Überwachung in der Produktion

Daniela Fisseler[1], René Reiners[2]

Abstract: Modelle im Industrie 4.0-Kontext sind ein wichtiges Hilfsmittel um Komplexität bei Produktionssystemen unter Kontrolle zu bekommen und sollen in der Betriebsphase verwendet werden um beispielsweise im Voraus mögliche Probleme zu erkennen. In dieser Arbeit stellen wir einen modellbasierten Ansatz zur Überwachung von Produktionsprozessen vor, bei dem Sensornetzwerke mit einem ausführbaren Modell der Produktion verknüpft werden. Ziel ist es, eine detaillierte Analyse auf Prozessschrittebene zu ermöglichen, etwa von Energieverbräuchen. Der vorgestellte Ansatz wird mit einem Beispiel in der Automobilproduktion verifiziert. Aus dem Modell zur Überwachung soll in einem nächsten Schritt ein Modell zur Simulation der Produktion erstellt werden.

Keywords: BPMN, Sensornetzwerk, Simulation, Geschäftsprozesse, Industrie 4.0

1 Einleitung

In Industrieunternehmen ist es häufig nicht möglich zu bestimmen, wie viele Ressourcen, in unserem Fall Energie, in einer einzelnen Station für einen einzelnen Artikel verbraucht wurden. Häufig sind die Sollwerte der Maschinenproduzenten für die einzelnen Maschinen bekannt sowie der Gesamtenergieverbrauch der Fabrik. Mit diesen Werten kann man aber nur den durchschnittlichen Energieverbrauch pro hergestelltem Artikel berechnen. Dabei ist eine prozessschritt- und stationsgenaue Datenerfassung und Auswertung aus verschiedenen Gesichtspunkten interessant. Eine genaue Datenerfassung durch Sensoren ermöglicht etwa eine vorausschauende Wartung, welche zu weniger und kürzeren Ausfällen in der Produktion führen kann. Außerdem werden mit diesen Daten neue Analysen ermöglicht: der Vergleich von verschiedenen Produkten oder verschiedene Produktionsstätten für gleiche Produkte. Die ermittelten Daten können dann zur Verbesserung der Ressourceneffizienz verwendet werden und dadurch für Firmen und Kunden monetäre und ökologische Vorteile bringen.

Industrie 4.0 beinhaltet die technische Integration von Cyber-Physischen-Systemen in die Produktion [Um13]. Sensornetzwerke und die Verarbeitung ihrer Daten spielen hierbei eine große Rolle. Außerdem sind Modelle in den Industrie 4.0-Konzepten ein wichtiges Instrument, um die Komplexität bei Produktionssystemen besser handhabbar zu machen. Es wird in der Regel zwischen Planungs- und Erklärungsmodellen

[1] Fraunhofer für angewandte Informationstechnik FIT, User-Centered Ubiquitous Computing, Schloss Birlinghoven, 53754 Sankt Augustin, daniela.fisseler@fit.fraunhofer.de
[2] Fraunhofer für angewandte Informationstechnik FIT, User-Centered Ubiquitous Computing, Schloss Birlinghoven, 53754 Sankt Augustin, rene.reiners@fit.fraunhofer.de

unterschieden. Erstere werden in der Planungsphase verwendet, um zu erklären, wie etwas umgesetzt werden soll. Erklärungsmodelle hingegen werden verwendet, um existierende Systeme zu erklären und Wissen über die Systeme zu erhalten, etwa durch Simulation, oder auch um Designentscheidungen zu verifizieren. Zukünftig sollen Erklärungsmodelle auch insbesondere in der Betriebsphase verwendet werden, um etwa „den regulären Betriebszustand zu überwachen, Verschleiß ohne Produktionsunterbrechung zu erkennen oder Komponentenausfälle und Störungen vorherzusagen“ [Um13].

Diese Industrie 4.0 Konzepte können genutzt werden, um eine prozessorientierte Überwachung umzusetzen. Diese Arbeit beschreibt einen Ansatz zur Modellierung von Produktionsprozessen und die erstellten Modelle mit Sensornetzwerken in Produktionsumgebungen zu kombinieren. Die Ausführung dieser Modelle erfolgt mit Hilfe einer Process Engine. Hierdurch wird eine prozessorientierte Überwachung der Produktion anhand von Sensorwerten und prozessrelevanten Ereignissen erreicht. So können Energieverbräuche zeitnah den Prozessschritten oder Subprozessen zugeordnet werden - eine nachgelagerte Zuordnung entfällt somit. Weiterhin können maschinelle Ausfälle vorausgesagt und frühzeitig Maßnahmen zur Behebung von Störungen eingeleitet werden. Durch die Verwendung einer graphischen Modellierungssprache werden die Prozessmodelle insgesamt leichter verständlich.

Ein vielversprechender Ansatz für die Modellierung ist Business Process Modelling and Notation (BPMN) 2.0., da es eine in der Industrie verbreitete, standardisierte, graphische Notation ist um Geschäftsmodelle darzustellen. Ein Ziel von BPMN ist die Bereitstellung einer Notation, die für Personen mit unterschiedlichen Rollen in einem Unternehmen einfach verständlich ist. Somit kann sie als Kommunikationsmittel eingesetzt werden, etwa zwischen technischen Entwicklern und Business Analysten. Der Standard definiert außerdem die XML Serialisierung von BPMN Modellen. Überdies hinaus gibt BPMN anderen XML-basierten Sprachen, wie WSBPEL, die für die Ausführung von Geschäftsprozessen konzipiert wurden, die Möglichkeit für eine graphische, geschäftsorientierte Visualisierung [Bu11].

Mit Hilfe dieses BPMN-Modells zur Überwachung wird ein Modell für die Simulation der Produktion erstellt. Dies findet Anwendung bei Tests oder Demonstrationen oder aber um Änderungen in der Produktion zu simulieren. Andere BPMN-Tools bieten zwar auch Simulationsmöglichkeiten für Prozesse, aber dort werden nur die Nachrichten simuliert, für die explizit Events definiert worden sind. Die Benutzung wäre aufwendig und unübersichtlich, da dann für jeden Sensor ein Event im Prozessmodell definiert werden muss, was in dem hier beschriebenen Ansatz nicht nötig ist. Eine Alternative wäre ein Simulator für Sensoren, der MQTT Nachrichten verschickt, jedoch müssten hierfür nochmals alle Sensoren neu modelliert bzw. konfiguriert werden.

Der restliche Aufbau des Artikels ist wie folgt: In Kapitel 2 werden verwandte Arbeiten vorgestellt; in Kapitel 3 wird erklärt, wie die Geschäftsprozessmodelle zur Überwachungen erstellt werden können inklusive einer Machbarkeitsstudie; Kapitel 4

erläutert beispielhaft, wie die Umwandlung in ein Simulationsmodell aussehen kann und Kapitel 5 beschreibt die nächsten forschungsrelevanten Schritte für den Ausbau des vorgestellten Ansatzes.

2 Verwandte Arbeiten

Bei der prozessorientierten Überwachung werden Sensornetzwerke mit der Ausführung von Geschäftsprozessen mit Hilfe von Ereignissen kombiniert. Es gibt Ansätze um Complex Event Processing und Prozess Management miteinander zu integrieren zu Event Driven Business Process Management. In [Am08] wird ein allgemeines Framework beschrieben, um dies zu erreichen. Die beiden unterschiedlichen Plattformen kommunizieren über Ereignisse, die entweder durch das IT-System ausgelöst werden oder die Prozess Engine. Um mögliche Probleme zu entdecken, die die korrekte Ausführung des Prozesses gefährden, werden Ereignis-Patterns definiert und die Ereignisse auf Übereinstimmung mit diesem geprüft. In [EH12] wird BPMN2.0 zum Modellieren der Hauptproduktionsprozesse verwendet. Eine Prozess Engine hilft Key Performance Indikatoren zu berechnen und komplexe Ereignisse zu erkennen und zu verarbeiten, so dass nur eine Plattform benötigt wird. In [Wa08] wird eine Architektur für ein einheitliches Ereignis Management bei komplexen Ereignissen vorgestellt.

In [Fr12] wird eine Erweiterung von BPMN vorgestellt um das Konzept von Business Activity Monitoring zu integrieren und Key Performance Indikatoren in BPMN darzustellen. In [Su13] wird eine Erweiterung von BPMN vorgestellt um eine Wireless Sensor Networks (WSN) in Prozessmodelle zu integrieren. Aus dem Modell soll dann Code generiert werden, um Daten aus dem WSN zu erfassen, zu aggregieren und Aktuatoren im WSN zu kontrollieren. In [Zo11] sollen Geschäftsprozessmodelle in der Produktion mit BPMN modelliert werden und BPMN wird hierfür erweitert um Materialflüsse, Maschinen Werkzeuge explizit modellieren zu können. BPMN kann in BPEL transformiert werden und in [Ko11] wird eine Klassifizierung und Evaluierung von verschiedenen BPEL Erweiterungen vorgenommen. In diesem Beitrag werden Alternativen zur Spracherweiterung untersucht, da diese potenziell die Komplexität erhöhen und somit sowohl das Modellieren als auch das Verständnis der Modelle erschweren.

Bei anderen Ansätzen geht es ebenfalls darum Sensordaten zu überwachen: In [WKN08] werden ausführbaren Workflows mit BPEL und Kontext-Information kombiniert, welche von Sensoren oder anderen Informationsquellen kommen, so dass direkt auf die vorhandene Situation reagiert werden kann und ggf. neue Prozesse etwa für die Bestellung von Ersatzteilen angestoßen werden können. In unserem Ansatz soll aber mit einer graphischen Sprache modelliert werden, weil diese auch für Nicht-Experten verständlich ist.

In [Am10] werden Prototypen für Demand-Side Energie Management beschrieben. In einem ersten Schritt wird analysiert, wieviel Energie benötigt wird, allerdings auf Ebene

eines einzelnen Produktstück oder auf Maschinenebene. Dafür müssen Sensoren und Smart Meter angeschlossen werden und die Verbräuche überwacht und analysiert werden. Danach soll dynamische Prozesskontrolle in einer Prozess Engine umgesetzt werden. In diesem Beitrag soll die Überwachung aber nicht auf Maschinen- oder Stückebene erfolgen sondern auf Prozessschritt- oder Subprozessebene.

3 Prozessorientierte Überwachung

3.1 Ziele

Prozessorientierte Überwachung bedeutet, dass die Werte, etwa Messwerte von Sensoren, die für die Überwachung relevant sind, bestimmten Prozessschritten und Prozessinstanzen zugeordnet werden und im Kontext des Prozesses betrachtet werden. Mit den entsprechenden Sensoren ermöglicht dies eine genaue Analyse der Stromverbräuche, z.B. für den Einbau einer Autotür bei der Automobilherstellung.

Prozessorientierte Überwachung in der Produktion kann mehrere Ziele verfolgen, zum Beispiel die Anwendungsfälle für Erklärungsmodelle in den Umsetzungsempfehlungen für den Industrie 4.0-Kontext [Um13]:

- Überwachung des regulären Betriebsstandes
- Erkennung von Verschleiß, etwa von Maschinenteilen, ohne den Produktionsprozess zu unterbrechen
- Vorhersage von Ausfällen oder Störungen von Komponenten

3.2 Anwendungsfall

Das Konzept der prozessorientierte Überwachung wurde anhand einer Automobilfertigungsstraße verifiziert. Abb. 1 zeigt einen Ausschnitt eines modellierten Schweißprozesses in der Automobilfertigung. Der Schweißprozess besteht aus 15 Stationen von denen 7 tatsächlich schweißen. In den anderen wird transportiert, ausgemessen oder gewartet. Es sollte eine Auswertung der Energieverbräuche pro Station erfolgen, welche hier als Subprozess modelliert wurden. Es wurden 220 Sensoren simuliert. Die Intermediate Catch Events warten auf ein Ereignis welches den Start einer einzelnen Station anzeigt. Dieses kann entweder durch Sensoren ausgelöst werden oder durch das Steuerungssystem. Das Ende der Bearbeitung in einer Station kann entweder implizit modelliert werden durch den Start der Bearbeitung in der nächsten Station oder, ähnlich wie beim Start, durch ein Ereignis ausgelöst durch einen Sensor oder das Steuerungssystem.

Um die Ereignisse zu empfangen, muss die Applikation sie vorher abonnieren. Dies

kann generisch geschehen, etwa durch die Benutzung von Wildcards.

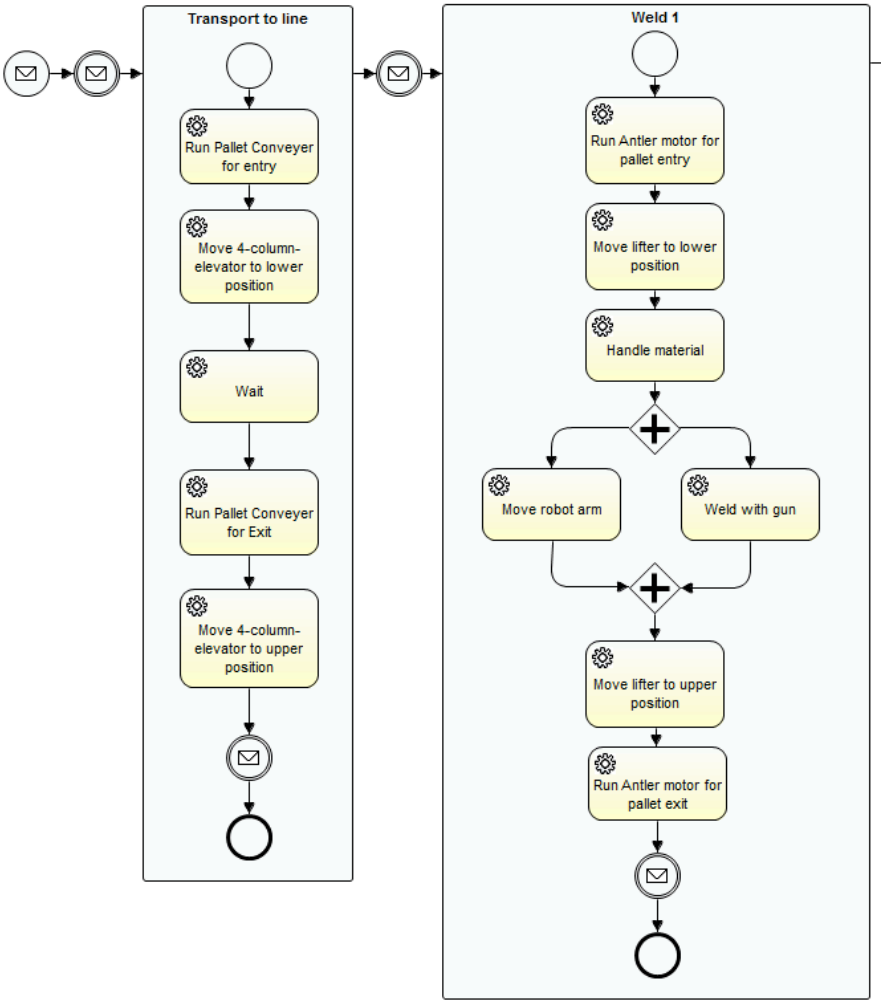


Abb. 1: Ausschnitt BPMN Diagramm Schweißen

Jede Automobilkarosse im Schweißprozess stößt eine neue Instanz an, die durch die Karosseriekennung identifiziert werden kann. Bei Eintreffen eines Sensorereignis wird geprüft, zu welcher Instanz dieses Ereignis gehört und ob es zu dem gerade aktiven Subprozess passt. Ist dies der Fall, werden Informationen über die Instanz und den Subprozess in den Body des Ereignisses geschrieben und je nach Bedarf erneut über den Broker verschickt (mit abgeändertem Topic um die Ereignisse von den ursprünglichen zu unterscheiden) oder direkt in eine Datenbank gespeichert, siehe auch Abb. 2.

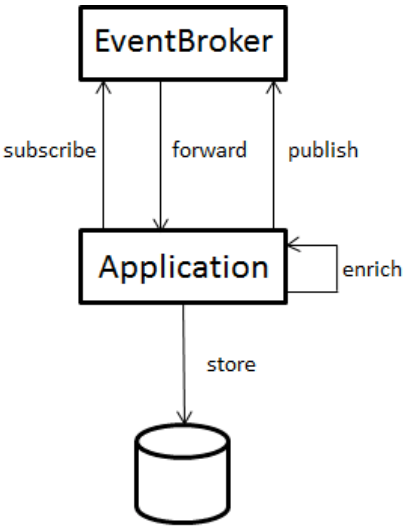


Abb. 2: Ereignisverarbeitung

3.3 Technische Umsetzung

Für die Modellierung und das Ausführen von Geschäftsmodellen wurde activiti[3] verwendet. Activiti ist ein Open-Source Workflow Management System, welches u.a. eine Process Engine und ein Plugin für Eclipse[4] beinhaltet, mit dem man Geschäftsprozesse modellieren kann. Das BPMN Modell selber kann sowohl in graphischer Form als auch als XML-Serialisierung bearbeitet werden.

Um Sensordaten mit Prozessschritten zu verknüpfen gibt es verschiedene Ansätze: man kann zum einen die Daten speichern und nachträglich den Prozessschritt und die Instanz anhand des Zeitstempels ermitteln oder man kann die Sensorevents zur Laufzeit mit den Prozessinformationen anreichern. In den hier beschriebenen Prototypen wird der erste Ansatz verwendet.

Ereignisse (Events) mit Sensordaten werden mittels eines Eventbrokers übertragen. Events bestehen aus Topics und einem Body. Jeder Sensor hat ein anderes Topic.

In einem ersten Prototyp wurden die Eventstopics manuell zugeordnet und zur Laufzeit aufgelöst. Schon bei mittleren Sensornetzwerken ist dies nicht mehr praktikabel und wird im nächsten Prototyp zumindest halb-automatisch geschehen. Die Zuordnung selber ist abhängig davon ob die Zuordnung auf Subprozess- oder Prozessschrittebene passieren soll. Die Zuordnung selber wird in der BPMN-Datei gespeichert. Um die

[3] http://www.activiti.org
[4] https://eclipse.org/ide

Komplexität nicht zu erhöhen, werden hier keine Erweiterungen an dem BPMN-Standard vorgenommen. Stattdessen werden die Mechanismen, die activiti bietet genutzt.

Im folgendem ist die XML-Serialisierung eines Prozessausschnitts dargestellt.

```
<process id="myProcess" name="My process"
isExecutable="true">
    <startEvent id="startevent1" name="Start">
    </startEvent>
    <endEvent id="endevent1" name="End"></endEvent>
    <serviceTask id="servicetask1" name="Service Task 1"
          activiti:class="de.beispiel.TaskHandler">
    </serviceTask>
    <sequenceFlow id="flow1" sourceRef="startevent1"
          targetRef="servicetask1">
    </sequenceFlow>
    <intermediateCatchEvent Id="msgintermcatch1"
          name="MessageCatchEvent">
          <messageEventDefinition messageRef="msg1Ref">
          </messageEventDefinition>
    </intermediateCatchEvent>
    <sequenceFlow id="flow4" sourceRef="servicetask1"
          targetRef=" msgintermcatch1">
    </sequenceFlow>
    <serviceTask id="servicetask2" name="Service Task 2"
          activiti:class="de.beispiel.TaskHandler">
     </serviceTask>
     <sequenceFlow id="flow5" sourceRef="msgintermcatch1"
          targetRef="servicetask2">
    </sequenceFlow>
     <sequenceFlow id="flow6" sourceRef="servicetask2"
          targetRef="endevent1">
    </sequenceFlow>
</process>
```

Für die Zuordnung der Topics bei serviceTasks wird das von Activiti implementierte „Field Injection“ verwendet. Es gibt zwei Arten der Injektion: für fixe String Werte und für Ausdrücke. Die implementierende Klasse muss diese Felder dann definieren und setter und getter-Methode hinzufügen. Die Zuordnung der Topics wird als JSON-Objekt in dem Modellierungswerkzeug als Field definiert oder direkt in die XML-Serialisierung eingebettet, was dann wie folgt aussieht:

```
<serviceTask id="servicetask1" name="Service Task 1"
    activiti:class="de.beispiel.TaskHandler">
    <extensionElements>
```

```
            <activiti:field name="testList">
                  <activiti:expression>
                  {
                  "sensors": [
                  {"topic": "RAW/MSRMT/L1/ST2/R5/S1"},
                  {"topic": "RAW/MSRMT/L1/ST2/R6/S1"},
                  {"topic": "RAW/MSRMT/L1/ST2/R7/S1"},
                  {"topic": "RAW/MSRMT/L1/ST2/R8/S1"}]
                  }
                  </activiti:expression>
            </activiti:field>
      </extensionElements>
  </serviceTask>
```

In diesem Fall werden dem ServiceTask1 vier verschiedene Topics und damit vier verschiedene Sensoren zugeordnet.

Für Message Catch Events wird ähnlich verfahren, nur dass die Injektion bei einem Listener erfolgt.

4 Simulation

Es gibt verschiedene Anwendungsfälle für eine Simulation des Prozesses: um die Implementierung zu testen oder das Prozessmodell zu prüfen, für Demonstrationszwecke oder auch um Änderungen im System zu simulieren und die Auswirkungen zu analysieren.

Um den Prozess zu simulieren wird neben dem Prozessmodell eine Simulation der Sensoren bzw. deren Nachrichten benötigt. Es gibt BPMN-Modellierungswerkzeuge, die die Simulationen für die erstellten Modelle ausführen können, allerdings würden in unserem Fall nur Nachrichten simuliert für die ein Catch-Event modelliert wurde. Die Nachrichten mit den zu überwachenden Sensoren würden nicht simuliert. Eine andere Möglichkeit wäre eine Simulation der Sensoren und der Versand über einen MQTT-Broker. Aber hier müssten alle Sensoren neu modelliert oder konfiguriert werden, welches bei einer mittleren Anzahl von Sensoren aufwendig ist.

Um den Aufwand möglichst gering zu halten, soll das Produktionsmodell, welches zur Überwachung verwendet wird auch für die Simulation verwendet werden. Dafür sind natürlich einige Änderungen nötig, da bei der Ausführung des Modells ja Nachrichten verschickt und nicht empfangen werden sollen. Außerdem sind zusätzliche Informationen notwendig, wie etwa die Bearbeitungsdauer in einer Station oder eines einzelnen Roboters.

Folgende Änderungen sollen vorgenommen werden:



Message Start Event: In activiti gibt es 4 verschiedene Arten von Start Events (die BPMN Spezifikation hingegen nennt 10 verschiedenen Events wenn man zwischen unterbrechenden und nicht-unterbrechenden Events unterscheidet [Bu11]): None Start Event, Message Start Event, Timer Start Event und Error Start Event. Abhängig von dem zu simulierenden Prozess, können verschiedene Ansätze nützlich sein. Wenn eine Produktionslinie simuliert werden soll, bei der regelmäßig eine neue Prozessinstanz gestartet werden soll, ist das Timer Event sinnvoll, ansonsten sollte a none start event verwendet werden. Um dies zu entscheiden, müssen Informationen über die Intervalllänge im Modell enthalten sein. Diese müssen manuell eingetragen werden. Nach dem StartEvent muss eine Nachricht verschickt werden, der im Modell zur Überwachung eine neue Instanz aktiviert. Dafür wird ein neuer ServiceTask generiert, der die Nachricht mit dem gemappten Topic verschickt.



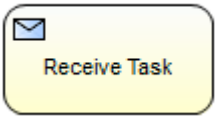
MessageIntermediateCatchEvent: dieses sollte ersetzt werden durch ein IntermediateThrowEvent, welches allerdings bisher nicht von Activi unterstützt wird, genauso wenig wie SendTasks für Nachrichten die an einen Message Broker versendet werden sollen. Deswegen sollten die IntermediateCatchEvents durch ServiceTasks ersetzt werden. Für ServiceTasks in activiti müssen Implementierungsklassen hinterlegt werden. In diesen Klassen kann das Versenden von Nachrichten zum MessageBroker implementiert werden.



IntermediateSignalCatchEvent: dies müsste genauer untersucht werden. Wahrscheinlich keine Ersetzung nötig, da Signale häufig im gleichen Prozess verschickt und empfangen werden (wenn von außerhalb verschickt werden soll, werden eher Nachrichten verwendet). Es ist aber möglich Signale sowohl programmatisch als auch zwischen Prozessen zu verschicken.

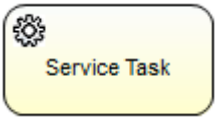


EventGateway: Normalerweise weist dieses Gateway darauf hin, dass es verschiedene Event-Möglichkeiten gibt und somit verschiedene Gateways folgen. Wie genau damit umgegangen werden soll, muss genauer untersucht werden.



ReceiveTask: sollte logisch gesehen durch einen SendTask

abgelöst werden. Laut einem activiti Entwickler[5] gibt es diesen bereits in einem experimentellen Stadium, ist aber noch nicht dokumentiert. Diesen SendTask kann man entweder wie einen MailTask oder wie einen WebserviceTask verwenden, bringt also keine neue Funktionalität. Insbesondere kann man keine Nachrichten an MessageBroker versenden, so dass auch dieser Task durch einen ServiceTask ersetzt werden muss.



ServiceTask: hier muss der Task nicht ersetzt werden, sondern nur die Implementierung. Wenn man davon ausgeht, dass ein Prozess mit einem Sensornetzwerk überwacht/ simuliert werden soll, welches etwa Energieverbräuche misst und dass der ServiceTask die Einheit repräsentiert, welches die Nachrichten verschickt, dann sollte die Simulationsimplementierung genau diese Nachrichten versenden.

Listener welche im Modell für die Überwachung definiert werden, sollten generell ersetzt werden (z.B. bei IntermediateCatchEvents) oder komplett entfernt (etwa bei ServiceTasks).

Das Modell braucht außerdem zusätzliche Informationen (zusätzlich zu den Informationen, wie häufig ein Prozess gestartet wird, falls dies regelmäßig und automatisch geschehen soll): MessageBroker URL und die Arbeitszeit der einzelnen Maschinen oder Maschinenteile.

5 Zusammenfassung und Ausblick

In diesen Artikel wurde ein Ansatz zur prozessorientierte Überwachung vorgestellt. Die Machbarkeit wurde anhand einer Produktionsstraße verifiziert. 220 Sensoren wurden simuliert und Prozessschritten zugeordnet. Mit Hilfe der Daten konnte eine genaue Analyse der Energieverbräuche auf Maschinen- und Prozessebene durchgeführt werden.

In den nächsten Schritten soll die Implementierung ausgebaut werden: der nächste Prototyp für die prozessorientierte Überwachung soll eine vereinfachte Zuordnung zwischen Prozessschritt und Ereignis Topic ermöglichen. Die Nützlichkeit der vorgestellten Kombination aus Prozess- und Simulationsmodell soll in einer weiteren Studie gezeigt werden.

Der Prototyp für die prozessorientierte Überwachung sollte auch noch auf Skalierbarkeit untersucht werden. Bei den bisherigen Tests hatte der Prozess maximal 130 Sensoren, die gleichzeitig Events verschickt haben bei 15 Prozessinstanzen. Das System war so konfiguriert, dass die Sensoren etwa einmal in der Sekunde ein Update bezüglich des

[5] http://forums.activiti.org/content/send-task-not-recognized-designer, 2015/01/05

Stromverbrauchs verschickt haben.

Es ist außerdem beabsichtigt, den Prototypen in einer realen Fabrikhalle zu installieren um dann evaluieren zu können, wie gut sich der Ansatz für die Überwachung des Produktionsprozesses und die vorausschauende Wartung eignet.

Danksagung

Diese Arbeit wurde von dem über das BMWI –geförderte Leitprojekt E3-Produktion[6] der Fraunhofer Gesellschaft unterstützt.

Literaturverzeichnis

[Am10] Ameling, Michael, et. al: "Demand-side Energy Management Systems for Manufacturing." In GI Jahrestagung (1), S. 467-472, 2010

[Am08] Ammon, Rainer von; et.al "Event-driven business process management and its practical application taking the example of DHL", In: 1st International workshop on Complex Event Processing for the Future Internet, Vienna Austria, S. 1-13, 2008

[Bu11] Business Process Model and Notation (BPMN) Version 2.0, 2011.

[EH12] Estruch, Antonio, Heredia Álvaro, José Antonio: "Event-driven manufacturing process management approach" In Business Process Management, Springer Berlin Heidelberg, S. 120-133, 2012.

[Fr12] Friedenstab, J.; et al.: "Extending BPMN for business activity monitoring." In 2012 45th Hawaii International Conference on System Science (HICSS), S.4158-4167, 2012.

[Ko11] Kopp, Oliver, et al.: "A classification of BPEL extensions." Journal of Systems Integration Vol 2, No 4 (2011), S.3-28. , 2011

[Su13] Sungur, C.T.; Spiess, P.; Oertel, N.; Kopp, O.: "Extending BPMN for Wireless Sensor Networks," in 2013 IEEE 15th Conference on Business Informatics (CBI), pp.109-116, 2013

[Um13] Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0, 2013.

[Wa08] Walzer, Karen, et al.: "Event-driven manufacturing: Unified management of primitive and complex events for manufacturing monitoring and control." in IEEE International Workshop on Factory Communication Systems, S. 383-391, 2008.

[WKN08] Wieland, Matthias; Kaczmarczyk, Peter, Nicklas, Daniela: "Context integration for smart workflows." In Pervasive Computing and Communications, PerCom 2008. Sixth Annual IEEE International Conference on, S. 239-242, 2008.

[Zo11] Zor, S.; Leymann, F.; Schumm, D.: “A proposal of BPMN extensions for the manufacturing domain,” in CIRP Conference on Manufacturing Systems, 2011.

[6] https://www.fraunhofer.de/de/forschungsfelder/fraunhofer-leitprojekte/e3-produktion.html

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Workshop zum Stand, den Herausforderungen und Impulsen des Geschäftsprozess-managements

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Workshop zum Stand, den Herausforderungen und Impulsen des Geschäftsprozessmanagements

Michael Fellmann[1], Agnes Koschmider[2], Ralf Laue[3], Andreas Schoknecht[2]

http://butler.aifb.kit.edu/ZuGPM/

Das Geschäftsprozessmanagement (GPM) ist mittlerweile eine etablierte Forschungsdisziplin. Um zukünftig die Relevanz der erreichten Forschungsergebnisse zu sichern und eine zielgerichtete Weiterentwicklung zu ermöglichen, müssen die Methoden und Instrumente des GPM an die sich im ständigen Wandel befindlichen Ansprüche aus Wissenschaft, Wirtschaft und Verwaltung angepasst werden. Darüber hinaus bedarf auch die Integration neuer Technologien wie beispielsweise RFID, mobile Geräte, Wearable Devices und Sensorik in das GPM der weiteren Untersuchung. Mit dem Workshop soll daher eine Plattform geschaffen werden, über die Ideen für die zukünftige Forschung im Kleinen, aber auch die Ausrichtung der GPM-Disziplin im Großen diskutiert werden können.

Der Workshop möchte somit dazu beitragen, wichtige Impulse und Anregungen für die zukünftige Forschung zu liefern. Beiträge zu diesem Workshop sollen auf bereits veröffentlichten Übersichtsartikeln (z.B. Literaturanalysen, Werkzeugvergleiche, Methodenvergleiche, Sprachvergleiche) basieren. Sie sollen eine Zusammenfassung des Artikels, eine Übersicht über Methoden und Ergebnisse sowie einen Diskussionsteil enthalten. Dabei dient die Übersicht über Methoden und Ergebnisse dazu, den Überblicksartikel standardisiert zu charakterisieren (bspw. nach Gegenstand des Überblicks, den angestrebten Erkenntnissen, den Adressaten der Forschung, den verwendeten Quellen und Methoden).

Zudem sollten Beiträge eine Zusammenfassung der wesentlichen Implikationen in Form von Thesen enthalten. Dem Diskussionscharakter des Workshops entsprechend, sollen die aus den Überblicksarbeiten abgeleiteten Thesen zwar im Einklang zu den Analyseergebnissen stehen, können jedoch deutlich über sie hinausgehen. D.h. Beiträge können auch weitreichende Interpretationen der Analyseergebnisse enthalten, die noch einer Validierung durch zukünftige Forschung bedürfen. Ausdrücklich erwünscht ist hierbei ein Bezug der Implikationen und Thesen zur Praxis oder die Interpretation der

[1] Universität Rostock, Institut für Informatik, Juniorprofessur Wirtschaftsinformatik, Albert-Einstein-Straße 22, 18059 Rostock, michael.fellmann@uni-rostock.de
[2] Karlsruher Institut für Technologie (KIT), Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB), Institut AIFB - Geb. 05.20, KIT-Campus Süd, 76128 Karlsruhe, agnes.koschmider@kit.edu / andreas.schoknecht@kit.edu
[3] Westsächsische Hochschule Zwickau, Fachgruppe Informatik, Dr.-Friedrichs-Ring 2a, 08056 Zwickau, ralf.laue@fh-zwickau.de

Befunde vor dem Hintergrund praktischer Erfahrungen.

Der Diskussionsteil sollte schließlich darauf eingehen, welcher Forschungsbedarf sich aus den Ergebnissen ableiten lässt, wie der Forschungsbedarf gedeckt werden kann, welche Folgen die Ergebnisse für die Wissenschaft und Praxis aufweisen oder welche neuen Anforderungen oder Anwendungsfälle sich für das Geschäftsprozessmanagement ergeben.

Zu diesem Workshop wurden fünf eingereichte Beiträge zu verschiedensten Aspekten des Geschäftsprozessmanagements angenommen. Die Beiträge geben dabei einen Überblick zur Modellierung wissensintensiver Prozesse, über den Einsatz verschiedener Medien zur Modellierung und die Möglichkeit zum Feedback in Geschäftsprozessen. Des Weiteren werden Übersichten zu Methoden zur Beherrschbarkeit von Variabilität in Geschäftsprozessmodellen oder der Transformation von Geschäftsprozessmodellen in Fähigkeiten-Modelle vorgestellt. Schließlich wird in einem eingeladenen Vortrag die Wiederverwendung von Geschäftsprozessmodellen thematisiert. Die wissenschaftlichen Beiträge werden durch eine Diskussionsrunde mit allen Teilnehmern abgerundet und bereichert.

Wir danken allen Einreichern für die sorgfältige Aufbereitung ihrer Arbeitsergebnisse sowie den Mitgliedern des Programmkomitees für die Mitwirkung bei der Begutachtung und Auswahl der Beiträge. Den Organisatoren der GI-Jahrestagung danken wir für die Unterstützung bei der Ausrichtung der Veranstaltung.

Programmkomitee: Patrick Delfmann (Universität Koblenz-Landau), Kathrin Figl (Wirtschaftsuniversität Wien), Hans-Georg Fill (Universität Wien), Oliver Kopp (Universität Stuttgart), Henrik Leopold (VU University Amsterdam), Jens Gulden (Universität Duisburg-Essen), Werner Schmidt (Technische Hochschule Ingolstadt), Meike Ullrich (Karlsruher Institut für Technologie)

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Modellierung wissensintensiver Geschäftsprozesse aus der Sicht des Wissensmanagements

Christoph Sigmanek[1], Birger Lantow[2]

Abstract: Bestehende Ansätze zur Modellierung wissensintensiver Geschäftsprozesse versuchen dem Charakter dieser Prozesse durch spezielle Konzepte und Methoden in der Modellierung gerecht zu werden. Dabei unterscheiden sich die Ansätze je nach Fokus der Modellierung erheblich. DeCo und KIPN zum Beispiel empfehlen eine Lockerung der Kontrollflussorientierung. KMDL erlaubt eine Modellierung hinunter bis auf die Ebene einzelner Personen. Auch SBPM und KPR setzen auf eine hohe Detaillierung und unterstreichen die Bedeutung des verteilten Modellierens. GPO-WM hingegen rät von zu vielen Details in den Prozessmodellen ab. Welcher Ansatz, welches Abstraktionsniveau ist nun aber passend für welche Modellierungsaufgaben aus Sicht des Wissensmanagements? Können die Modelle auch für andere Aufgaben wiederverwendet werden? Die Suche nach dem „richtigen" Weg zur Modellierung wissensintensiver Prozesse und sich daraus ableitende Fragestellungen stehen im Fokus der Diskussion.

Keywords: Wissensmanagement, Geschäftsprozessorientiertes Wissensmanagement, Wissensintensive Prozesse, Prozessmodellierung

1 Einleitung

Wissen ist heute als eine wichtige Unternehmensressource anerkannt. Daraus leitet sich das Wissensmanagement als Managementaufgabe ab. Geschäftsprozessorientiertes Wissensmanagement zielt dabei auf den Umgang mit Wissen und Anforderungen an Wissen und Wissensaktivitäten (Wissen nutzen, erzeugen, transferieren) in Geschäftsprozessen ab. Remus setzt wissensintensive Geschäftsprozesse in den Fokus eines prozessorientierten Wissensmanagements [Re02, S. 108]. Hier liegen die größten Erfolgspotentiale für das Wissensmanagement. Tabelle 1 fasst die typischen Merkmalsausprägungen wissensintensiver Prozesse zusammen. Sie kommen häufig in wissensintensiven Domänen vor und sind durch eine hohe Komplexität gekennzeichnet. Ihr Ablauf variiert stark, sodass ein hoher Koordinations- und Kommunikationsaufwand erforderlich ist. Wissensintensive Prozesse weisen häufig eine schwache Struktur mit einer hohen Anzahl an Beteiligten (Experten) auf und sind schwer planbar. Aufgrund ihrer Aufgabenbeschaffenheit sind die Aufgaben schwer auf andere Bearbeiter übertragbar [Re02, S. 104-117]. Heisig sieht als maßgebendes Kriterium von wissensintensiven Prozessen, dass sie „sich hinsichtlich ihrer benötigten Wissensinhalte

[1] Universität Rostock, Institut für Informatik, Lehrstuhl für Wirtschaftsinformatik, Albert-Einstein-Straße 22, 18059 Rostock, win.office@uni-rostock.de
[2] Universität Rostock, Institut für Informatik, Lehrstuhl für Wirtschaftsinformatik, Albert-Einstein-Straße 22, 18059 Rostock, birger.lantow@uni-rostock.de

Merkmalsklasse	Typische Merkmalsausprägungen für wissensintensive Geschäftsprozesse
Prozessübergreifende Merkmale *(Wissensintensive Domäne)*	☐ Oftmals dezentrale, stark vernetzte Organisation mit gezielter Förderung des Wissensaustauschs z. B. durch wissensorientierte Anreizsysteme ☐ Wissensintensive Branche (Schlüsseltechnologien) ☐ Komplexe Beziehungen zu anderen Prozessen
Prozessbezogene Merkmale *(Durch Unbestimmtheit und Kontext-/Wissensabhängige Variabilität gekennzeichnet)*	☐ Komplexe Prozesse mit einer Vielzahl voneinander stark abhängiger Einzelaufgaben, Personen, die in interdisziplinären Teams zusammenarbeiten ☐ Viele Sonderfälle und unvorhersehbarer Ablauf und unbestimmtes Ergebnis ☐ Schwach strukturiert, können erst rückwirkend modelliert werden ☐ Hoher Koordinierungs- und Kommunikationsaufwand zwischen Prozessbeteiligten, erfordert Expertise verschiedener Fachrichtungen ☐ Es entstehen wissensintensive Produkte und Dienstleistungen ☐ Nur ungenaues Controlling möglich, häufig qualitative Ziele ☐ Prozess hat wenig Durchführungen mit langer Dauer ☐ Prozess ist einzelfallbezogen, sachbezogen, kein Routineprozess
Aufgabenbezogene Merkmale *(schwer übertragbar)*	☐ Produktivität der Wissensarbeit meist nicht messbar ☐ Lange Lern- und Übungszeiten nötig ☐ Arbeitsplatz ist chaotisch, es herrscht Unordnung ☐ Aufgaben sind kommunikationsorientiert, informationslastig, argumentationsbasiert, einzelfall- oder sachbezogen ☐ Typisch: Entscheidungsaufgaben, Problemlösungsaufgaben, Analyse- und Bewertungsaufgaben, Führungs- u. Steuerungsaufgaben
Mitarbeiterbezogene Merkmale *(Experten)*	☐ Hohe Mitarbeiterautonomie und somit großer Einfluss auf das Ergebnis ☐ Unstrukturierte und individualisierte Arbeitsregeln und Routinen ☐ Hohe Kompetenz, Lernfähigkeit, Kreativität und Innovation gefordert
Ressourcenbezogene Merkmale *(Komplexe Ressourcen)*	☐ Instrumente des Wissensmanagements werden viel eingesetzt, Wissensaustausch erfolgt informell, „face to face“ oder dokumentenorientiert ☐ Wissen oft schwer zugänglich und stark kontextabhängig ☐ Hoher Wissensumschlag, nötige Expertise auf dem Markt ist kostenintensiv

Tab 1.: Merkmale wissensintensiver Geschäftsprozesse nach Remus [Re02]

nicht, oder nur begrenzt vorausplanen lassen“[He02]. Um die Wissensunterstützung von Prozessen modellieren zu können, ist es folglich nicht mehr ausreichend, den Prozess auf eine Folge von Aktivitäten, Ereignissen und Entscheidungen zu beschränken. Vielmehr

ist es notwendig, dass aus Sicht des Wissensmanagements bedeutende Komponenten dargestellt werden können und dass die Modellierungsmethodik den speziellen Eigenschaften dieser Prozesse gerecht wird.

Zu berücksichtigende Modellkomponenten sind:

1. Wissensaktivitäten
   a) Wissenserzeugung und Wissensnutzung [Al98, S. 165f.]
   b) Wissenstransfers
2. Wissensressourcen
   a) Wissensträger [Al98, S. 165f.]
   b) Wissensquellen
   c) Wissensstruktur
3. Rahmenbedingungen
   a) IuKT-Involvierung, eingesetzte Technologien [Sc98, S.63-65]
   b) Organisatorische Anforderungen und Unternehmenskultur [Le07]

Gleichzeitig muss die hohe Komplexität und Variabilität der wissensintensiven Prozesse durch die Modellierungsmethodik berücksichtigt werden.

Im Geschäftsprozess werden Aktivitäten in einer logischen Abfolge von Personen, die in Rollen agieren, ausgeführt. Zur erfolgreichen Bearbeitung von wissensintensiven Aufgaben müssen die Bearbeiter ihren Wissensbedarf unter Zuhilfenahme von Wissensaktivitäten decken. Dabei interagieren sie mehr oder weniger effizient über verschiedene Medien mit anderen Personen und IT-Systemen. Bei diesen Interaktionen wird ständig neues Wissen erzeugt, welches aber nur bedingt erhalten bleibt. Um diese Prozesse in ein Modell zu überführen, ist eine strukturierte Vorgehensweise nötig, die ein richtiges Detailniveau wählt. Da gute Modelle einen zweckmäßigen Ausschnitt der Realwelt darstellen, ist es möglich, über eine Analyse der Modelle Schlüsse zu ziehen, wie Prozesse in der Realwelt verbessert werden können. Weiterhin bietet die Modellierung den Vorteil, dass Wissen über Geschäftsprozesse dokumentiert oder geteilt werden kann. Als Voraussetzung für die Erstellung und Auswertung eines korrekten Modells impliziten und expliziten Wissens werden im folgenden Abschnitt State-of-the-Art-Methoden zur Erfassung und Analyse wissensintensiver Geschäftsprozesse vorgestellt. Im dritten Abschnitt folgt dann eine Evaluierung der vorgestellten Ansätze in Bezug auf die eingangs formulierten Anforderungen an die Modellierung. Der letzte Abschnitt fasst die gefundenen offenen Herausforderungen für die Modellierung wissensintensiver Prozesse nochmals zusammen.

2 Ansätze zur Modellierung und Analyse von wissensintensiven Geschäftsprozessen

Es gibt in der Literatur eine Vielzahl von Ansätzen zur Modellierung wissensintensiver Prozesse bzw. Modellierung von Aspekten des Wissensmanagements in Prozessmodellen. Dieser Abschnitt stellt eine Auswahl daraus vor. Es sind Ansätze dabei, die basierend auf einer Auswertung der Zitationshäufigkeit (>50 Zitate) mit Google-Scholar einen hohen wissenschaftlichen Impact haben. Ausnahmen bilden DeCo und KIPN (>10 Zitate), die neueren Datums sind.

2.1 Knowledge Modeling and Description Language

Die Knowledge Modeling and Description Language (KMDL) ist eine Methode zur Modellierung wissensintensiver Geschäftsprozesse, die zum Zeitpunkt der Arbeit noch aktiv weiterentwickelt wird [Gr09, S. 76-79] [KM14a]. KMDL grenzt sich durch eine Personen- bzw. Instanz-bezogene Informations- und Wissensmodellierung von anderen Ansätzen ab. Die Methode sieht dabei eine explizite Modellierung einzelner Wissenskonversionen nach Nonaka und Takeuchi [NT95] vor, aus der anschließend formale Reporte erstellt werden können. Zusätzlich stellt die Methode verschiedene Analysesichten und Vergleichspattern zur Analyse bereit.

Der KMDL-Methode liegt ein neun-Phasen Vorgehensmodell zugrunde, welches für erfolgreiche Projekte die aktive Teilnahme des Projektpartners voraussetzt. Zunächst werden in Phase 0 organisatorische Rahmenbedingungen gesetzt. In Phase 1 folgt die Definition von angestrebten Zielen des Projekts. Anschließend, in Phase 2, werden die Prozesse beim Projektpartner iterativ erfasst, verfeinert und validiert. Daraus werden in Phase 3 wissensintensive Prozesse abgeleitet. In Phase 4 werden diese wissensintensiven Prozesse iterativ modelliert und in Phase 5 auf Verbesserungsmöglichkeiten sequenziell analysiert. Konkret werden in der Analyse zunächst Schwachstellen analysiert, um daraus dann Verbesserungsvorschläge abzuleiten. Danach werden diese klassifiziert und bewertet. Abschließend kommt es zu einer Evaluation der Verbesserungspotenziale. In der anknüpfenden Phase 6 wird jetzt mit dem Partner ein Sollkonzept entwickelt, welches in Phase 7 umgesetzt wird. In der letzten Phase 8 wird nun der gesamte Prozess mit dem Projektpartner evaluiert. [Gr09, S.75f]

Für KMDL gibt es derzeit drei Sichten, mit denen unterschiedliche Anforderungen der Modellierung wissensintensiver Prozesse abgedeckt werden. Die Prozesssicht zeigt die Geschäftsabläufe auf einer abstrakten Ebene. Dort werden einzelne Aufgaben in ihrer logischen Reihenfolge mit beteiligten Ressourcen dargestellt. Aufgaben können in der Aktivitätssicht detailliert nach ablaufenden Wissensumwandlungen (z.B. Sozialisierungen) aufgeschlüsselt werden. Die Aktivitätssicht ist auch Grundlage für die Kommunikationssicht, in der beschrieben wird, wie einzelne Wissenstransformationen in Konversationen ablaufen. Dabei werden Konversationen nach Ort (gleicher Ort/verschiedener Ort) und Zeit (synchron/asynchron) unterschieden. [Po09, S. 21-45]

Sobald die Modelle erfasst und validiert sind, beginnt in KMDL die Analyse. In der Analyse werden anfangs Objekthäufigkeiten und Häufigkeiten der Konversationsarten (z. B. Sozialisierungen) gezählt und entsprechend ausgewertet. Eine häufig vorkommende Sozialisierung kann beispielsweise andeuten, dass zu wenig Wissen expliziert wird. Eine immer wiederkehrende Wissensressource oder eine Person, die an vielen Stellen beteiligt ist, weist analog dazu auf einen eventuellen Flaschenhals oder eine Schlüsselfunktion hin. Dementsprechend werden Wissensbedarfe mit Wissensangeboten abgeglichen und es findet eine Untersuchung der Modelle auf bestimmte Muster statt, für die es konkrete Handlungshinweise gibt. [Po09, S. 49-79]

KMDL bietet einen ganzheitlichen Ansatz zur Modellierung und Verbesserung von wissensintensiven Prozessen. Für die Modellierung mit KMDL steht das Tool K-Modeler [KM14b] zur Verfügung. Kritisiert wird die Methode für den zusätzlichen Aufwand, der beim Erheben der einzelnen Wissensinformationen entsteht, welcher nur durch eine bessere Schließung auf effektivere Wissensmanagement-Maßnahmen gerechtfertigt werden kann [KBL13, S. 417]. Damit ist diese Methode sehr zeitaufwendig und die Ergebnisse hängen maßgeblich von dem Vertrauenslevel der Interviewten und den Fähigkeiten des Interviewers ab [MBG12, S.362f].

2.2 Knowledge Process Reengineering

Der Ansatz des Knowledge Process Reengineering (KPR) [Al98, S.163-168] ist ein Sieben-Phasen-Konzept zur Verbesserung des Umgangs mit der Ressource „Wissen". Insbesondere zielt der Ansatz auf effektiven Wissensaustausch, gute Dokumentationen und leichten Zugang zum Wissen. KPR wurde für den Einsatz in Unternehmen entwickelt und kann durch ARIS-Modelle unterstützt werden. Die einzelnen Phasen, beginnend mit der strategischen Wissensplanung, weitergehend über die Ist-Analyse und Soll-Konzeption, bis hin zur Umsetzung, laufen in KPR linear ab, d.h. es sind keine Schritte zurück in eine bereits abgeschlossene Phase vorgesehen. Stattdessen sieht der Ansatz vor, dass in der Endphase ein fortlaufender Test- und Verbesserungsprozess angestoßen wird.

KPR beginnt mit einer Strategischen Wissensplanung, bei der entschieden wird, wie Wissensmanagement die strategischen Unternehmensziele unterstützen kann. Als Hilfe dienen dabei Modelle, welche die Kerngeschäftsprozesse des Unternehmens darstellen und diese mit den strategischen Unternehmenszielen in Beziehung setzen. In der anschließenden Ist-Erhebung der Wissensverarbeitung müssen die Geschäftsprozesse als EPK-Form vorliegen (da enge Kopplung an ARIS). Dann müssen Wissensträger, Wissenskategorien und Wissensbedarfe in Wissensstrukturdiagrammen, Wissenslandkarten und zusätzliche Informationen in den EPK-Diagrammen erfasst werden. [Al98, S.164-166]

Nachdem der Ist-Zustand erhoben wurde, beginnt die Analyse der Ist-Situation. Hier werden kritische Wissensmonopole, nicht befriedigte Wissensbedarfe, ungeeignete Wissensprofile von Mitarbeitern etc. aufgedeckt. Die darauf aufbauende Soll-

Konzeption der Wissensverarbeitung entwirft jetzt Maßnahmen, um die zuvor aufgedeckten Schwachstellen zu bereinigen, indem beispielsweise Soll-Wissensprofile für Mitarbeiter oder Änderungen der Geschäftsprozesse definiert werden. Nachdem das Soll-Konzept feststeht, werden im Realisierungskonzept „Organisation“ Mitarbeiter-schulungen für den Umgang mit den geänderten Prozessen und neuen Informations-technologien entwickelt, während im Realisierungskonzept „Informations-technik“ parallel dazu geeignete IuK-Unterstützungen (z. B. Intranetlösungen) entwickelt, Inhaltsstrukturen festgelegt und Integrationen verschiedener Systeme realisiert. In der Umsetzung der Realisierungskonzepte werden die entwickelten Soll-Maßnahmen durchgeführt und nach der Einführung getestet, sowie ggf. verbessert. [Al98, S.166-168]

KPR bietet damit einen Ansatz, der darauf abzielt, Technologien aus dem Wissensmanagement in den Arbeitsabläufen der Mitarbeiter zu verankern. Die starke Abhängigkeit von der zugrundeliegenden ARIS-Architektur, die Forderung das gesamte Wissen eines Unternehmens zu modellieren, und die fehlende Konkretisierung einzelner Arbeitsschritte sowie Weiterentwicklung der Methode führen dazu, dass KPR nur ein spezielles Vorgehensmodell zur Einbindung von IT in wissensintensive Geschäfts-prozesse darstellt.

2.3 PROMOTE

Hinkelmann et al. [HKT02, S. 65-68] haben mit process-oriented methods and tools for knowledge management (PROMOTE) eine technologieunabhängige Methode zum Management von Funktions- und Prozesswissen vorgestellt. Es ist eine Weiterentwicklung des Business Process Management Systems Rahmenwerks (BPMS) [Ka95] und ergänzt dieses durch das Softwarewerkzeug PROMOTE [BO14]. PROMOTE konzentriert sich auf die Identifikation, Modellierung und Integration von Prozessen, die Wissen suchen und generieren. Die Software unterstützt dabei die Bearbeitung wissensintensiver Aktivitäten, indem Wissensprozesse kontextspezifisch aktiviert werden können. Weiterhin bietet sie Wissenslandkarten und Topic Maps als konfigurierbare Wissensmanagement-Instrumente und schließlich stellt PROMOTE Managementfunktionen für Wissensflüsse und eine modellbasierte Indizierung von Dokumenten mit prozess- und rollenspezifischen Zugriffsrechten bereit.

Als Voraussetzung für die Methode gelten die Annahmen, dass a) Wissensprozesse genauso wie Geschäftsprozesse modelliert werden können und b) Aktivitäten in Geschäftsprozessen Wissen nutzen. Basis für den Einsatz von PROMOTE sind die Prozessschritte, die dem Bearbeiter ein hohes Maß an Gestaltungsfreiraum zusprechen und je nach Bearbeiter und Wissensstand in Verlauf und Ergebnis variieren. Die allgemeine Zielstellung ist dabei eine Katalyse der Wissensflüsse zwischen wissensintensiven Geschäftsprozessen. Diese Wissensflüsse können innerhalb eines Geschäftsprozesses, innerhalb eines Projektes geschäftsprozessübergreifend und sogar projektübergreifend auftreten. Außerdem sind Wissenszuflüsse von externer Seite durch Schulungen, Internetrecherche u.Ä. möglich. [HKT02, S. 68-71]

Die Realisierung erfolgt in den fünf Phasen Aware Enterprise Knowledge, Discover Knowledge Process, Modelling Knowledge Processes and Organisational Memory, Making Knowledge Processes and Organisational Memory operational und Evaluate Enterprise Knowledge. In der ersten Phase werden Unternehmensziele angepasst und strategisch festgelegt. Das sind z. B. Produkte, Dienstleistungen, finanzielle Vorgaben und die Entwicklung der Kernkompetenzen. Ziel ist dabei, dass Geschäftsstrategie und Wissensstrategie in Einklang gebracht werden. [HKT02, S. 73-76]

In der anschließenden Discover-Knowledge-Process-Phase wird Prozesswissen, d. h. Wissen über die logische Abfolge der Aktivitäten innerhalb eines Prozesses, samt daran beteiligter Personen, Organisationseinheiten, Anwendungssysteme und Ressourcen modelliert. Zusätzlich wird Wissen mit hohem Wirkungspotenzial durch Experten des Fachs festgehalten. Dazu gehören beispielsweise entscheidungskritisches Wissen und Wissen zum Erstellen einer Dienstleistung oder eines Produktes (Funktionswissen). Für die wissensintensiven Aktivitäten werden außerdem Arten des darin verarbeiteten Wissens, Wissensträger, Wissensflüsse und die Formen, in denen das jeweilige Wissen vorliegt, erfasst. Nach der Modellierung der Geschäftsprozesse finden Modellierung und Zuordnung der Wissensprozesse in der dritten Phase statt. Wissensprozesse sollen Wissensflüsse ersetzen, indem sie dem Wissensfluss eine Methodik geben. Wenn ein Bearbeiter Wissen für die Erfüllung seiner Aufgabe benötigt, dann stehen ihm bestimmte Möglichkeiten zur Verfügung, um an dieses Wissen zu kommen. Beispielsweise kann er sich an seinen Kollegen wenden oder in Gelben Seiten nach einem Experten für seinen Wissensbedarf suchen. Um Dokumente später wiederauffindbar und somit nutzbar zu machen, werden diese mit Metainformationen bereichert. Ein Dokument bekommt dazu Änderungsdatum, inhaltliche Stichworte (Tags) aus Folksonomien (Sammlungen von Tags), Autor und weitere Informationen, die idealerweise durch die entsprechenden Wissensstrukturen bereits vorgegeben werden und so den technischen Teil des organisatorischen Gedächtnisses bilden. [HKT02, S. 76-84]

Phase 4, Making Knowledge Processes and Organisational Memory operational, implementiert diese Wissensprozesse in bestehende Software, sodass ein Bearbeiter während seiner Arbeit dann sofort sehen kann, welche Möglichkeiten ihm zur Befriedigung seines Wissensbedarfs zur Verfügung stehen. Durch Verlinkungen können dann z. B. die Gelben Seiten zur Findung eines Experten mit einem Klick aufgerufen werden, wobei bestimmte Suchparameter bereits vorher ausgefüllt wurden. Eine Evaluierung des Einsatzes von PROMOTE findet dann in der 5. Phase statt. So kann der Beitrag des Wissensmanagements am Unternehmenserfolg gemessen werden. [HKT02, S. 84-90]

2.4 Declarative Configurable

Declarative Configurable (DeCo) ist eine Kombination aus deklarativer Modellierung, Modellverifikation und Variabilitätsmodellierung zur Erfassung wissensintensiver Prozesse. Die wissensintensiven Prozesse werden in DeCo auf drei Ebenen modelliert.

Die Abstrakteste Ebene ist at design, in der eine konfigurierbare, auf die Prozessziele ausgerichtete, nichtdeterministische Spezifikation erstellt wird. In der at-deployment-Ebene wird der Prozess in einen Kontext konfiguriert, der dem Einsatzgebiet nahe kommt. Schließlich entsteht eine voll deterministische Prozessspur at execution, die einen einzelnen Prozess abbildet. [RN11, S. 1-2]

Geschäftsprozesse werden dabei in präskriptive Prozesse und deskriptive Prozesse unterteilt. Präskriptive Prozesse haben vorhersehbare Abläufe einfacher Aufgaben und können zur Designzeit vollständig spezifiziert werden. Den Gegenpol bilden die deskriptiven Prozesse, zu denen auch die wissensintensiven Prozesse gehören. Diese komplexen Aufgaben, die auf Zusammenarbeit verschiedener Akteure beruhen, lassen sich zur Designzeit nur skizzieren. Prinzipien, die DeCo zugrunde liegen, sind: „Very little is certain at design-time“ und „Fixed constraint often means lost opportunities“. Innerhalb der at-design-Ebene wird deshalb für deskriptive Prozesse noch kein Kontrollfluss gefordert. So bleibt die Konfigurierbarkeit unlimitiert und kritische Entscheidungen können auf einen späteren Zeitpunkt hinausgezögert werden. [RN11, S. 2-5]

Um zur Anwendung zu kommen, werden Prozesse in einen spezifischen Kontext in der at-deployment-Ebene konfiguriert, wobei einige Details wegen ihrer Unbestimmtheit noch nicht vorkonfiguriert werden können. Bei konfigurierbaren Prozessen werden z. B. Aufgaben Rollen zugewiesen, Aufgaben angeordnet oder geltende Regeln ausgewählt. In der at-execution-Ebene werden die vorkonfigurierten Prozesse schließlich ausgeführt und hinterlassen Prozessspuren, die abgespeichert werden und so zur Konstruktion einer Wissensbasis beitragen und damit auch zur Verbesserung künftiger Prozesse beisteuern. [RN11, S. 2-9]

DeCo hilft damit beim kontrollierten Zusammenbau bedeutender Prozessspezifikationen aus vordefinierten Prozessteilen oder Prozessvarianten. Die Designphase wird durch Leitfragen gesteuert und nach jeder Ausführung werden neue mögliche Pfade in das initiale oder angepasste Modell eingearbeitet. Realisiert wird DeCo durch eine Anpassung des BPMN-Standards: Optionale Objekte werden durch gestrichelte Linien gekennzeichnet, konfigurierbare durch fette Linien. Weiterhin werden Objekte durch Tags (z. B. <IN> für eingehende Informationen) zur Beschreibung wissensintensiver Prozesse angereichert. Hauptsächlich wird durch diesen Ansatz die Variabilität von wissensintensiven Prozessen abgedeckt. Für künftige Verbesserungen wird die Auswertung der Prozessspuren empfohlen. [RN11, S. 5-10]

2.5 GPO-WM

Heisig [He02, S. 47-59] zeigt mit „Geschäftsprozessorientiertes Wissensmanagement“ (GPO-WM) ein Acht-Phasen-Vorgehensmodell zur Einführung von Wissensmanagement. Weiterhin können damit Stärken und Potenziale bezüglich des Umgangs mit der Ressource „Wissen“ im Unternehmen bestimmt werden. Wichtige Quintessenzen bei GPO-WM sind, dass der Detaillierungsgrad der Modellierung nicht

übertrieben wird, und dass Nähe zum Unternehmen gewahrt wird, indem unter anderem die Modellierung in einer unternehmensspezifischen Sprache durchgeführt wird. Zur Abgrenzung der problematischen Aktivitäten wird die Leitfrage: Enthält die Aktivität Basisaufgaben des Wissensmanagements? vorgeschlagen. Basisaufgaben sind dabei Wissen erzeugen, Wissen speichern, Wissen verteilen und Wissen anwenden.

Innerhalb der Analyse steht nicht die Optimierung einzelner Aktivitäten wie das Abspeichern expliziten Wissens in einer Datenbank im Vordergrund, sondern vielmehr eine Berücksichtigung des gesamten Rahmens, d. h. Fragestellungen wie: Wo wird das hier erzeugte Wissen wiederverwendet? stehen im Fokus. Probleme werden anhand von Leitfragen identifiziert und gegebenenfalls durch Best-Practice-Lösungen des Wissensmanagements ersetzt. So können auch Probleme entdeckt werden, die nicht in einem Modell abgebildet sind. Als Ergebnis werden die Wissensmanagementbausteine implementiert und in die jeweiligen Geschäftsprozesse integriert. [He02, S. 59-64]

2.6 KIPN

França et al. [FBS12] bemerken, dass es bereits zahlreiche Methoden zur Modellierung von wissensintensiven Prozessen gibt und prüfen, inwieweit diese wissensintensive Prozesse abbilden können. Wie auch Gronau [Gr09, S. 69-71] in einer ähnlichen Studie kommen sie zu dem Schluss, dass kein Ansatz aus der Literatur alle relevanten Bereiche abdeckt. Sie gehen einen Schritt weiter und untersuchen auch bereits etablierte Modellierungssprachen wie EPKs und BPMN auf die Kriterien. Deutlich wird, dass EPKs und BPMN bereits viele der durch Remus [Re02, S. 115-116] definierten Anforderungen erfüllen. Defizite bestehen in der Darstellung schwach strukturierter Prozesse, der Beziehungen zu anderen Geschäftsprozessen, des Wissensaustauschs und der kurzen Halbwertszeit des Wissens.

Als Ergebnis stellen sie eine Ontologie vor [FBS12, S. 499-504], auf deren Grundlage die Knowledge Intensive Process Notation (KIPN) entwickelt wurde. KIPN ist eine graphische Notation, die aus fünf Diagrammen zusammengesetzt ist. In dem KIP-Diagramm werden Aktivitäten, samt Geschäftsregeln, Verbindungen und Abstraktionslevel dargestellt. Der Kontrollfluss einzelner Aktivitäten muss in KIP-Diagrammen nicht zwangsweise vorgegeben werden. Kommunikationen zwischen den Beteiligten, d. h. ausgetauschte Nachrichten, Wissensbeschaffung und Wissensteilung, werden in einem Sozialisierungsdiagramm dargestellt. In einem Entscheidungsdiagramm werden schließlich Alternativen und deren Vor- und Nachteile aufgelistet. Außerdem stellt die Notation Diagramme zur Ziel- bzw. Rollenmodellierung bereit [FBS13].

3 Bewertung der vorgestellten Ansätze

Im vorhergehenden Abschnitt wurden unterschiedliche Ansätze zum Umgang mit wissensintensiven Geschäftsprozessen vorgestellt. In den einzelnen Ansätzen wird

deutlich, dass wissensintensive Prozesse aufgrund ihrer Beschaffenheit anders als normale Prozesse behandelt werden müssen. Sämtliche Methoden heben hervor, dass eine richtige Fokussierung bei der Modellierung für den Output einer Analyse maßgebend ist. DeCo und KIPN empfehlen eine Lockerung der Kontrollflussorientierung vieler Modellierungssprachen. Zu Beginn einer Modellierung oder Analyse steht immer ein Abgleich mit den Wissens- bzw. Unternehmenszielen. KMDL bietet die Möglichkeit, auf Individuen-Ebene zu modellieren und setzt auf eine starke Einbeziehung des Modellierten. Dadurch wird ein exaktes Replikat der Realität geschaffen, das u. U. nicht übertragbar ist und aufgrund der in DeCo beschriebenen Variabilität nur eingeschränkte Wartbarkeit aufweist.

KPR empfiehlt eine verteilte Modellierung. Dadurch, dass Konzepte und nebenläufige Aktivitäten an anderer Stelle spezifiziert sind, kann eine semantische Gleichheit im ganzen Diagramm garantiert werden. Sowohl bei verteilter Modellierung als auch bei einer Modellierung in einem zentralen Modell wird eine Modellierung in Phasen vorgeschlagen. In einigen Ansätzen hat die detailgetreue Modellierung für eine anschließende Analyse einen hohen Stellenwert, während GPO-WM von einer zu hohen Detaillierung abrät. Als Lösungen sehen die meisten Methoden vor, dass konkrete Wissensmanagementsoftware eingepflegt und in die Geschäftsprozesse integriert wird. GPO-WM stellt dazu Best Practices bereit, die bestimmte Kompetenzen abdecken können. Durch die Vielfalt an Zielstellungen existiert eine Vielzahl an unterschiedlichen Modellierungssprachen, von denen sich bis heute jedoch keine in der Literatur durchsetzen konnte.

Tabelle 2 stellt die wesentlichen Merkmale der einzelnen Ansätze überblicksmäßig dar. Dabei werden zunächst generelle Modellierungsaspekte wie Zielstellung, Methodische Unterstützung und Werkzeugbindung aufgegriffen. Weiterhin geht die Tabelle auf die spezifischen Anforderungen der Modellierung wissensintensiver Prozesse aus Abschnitt 1 ein. Bei den Wissensaktivitäten wird zwischen der Darstellung von Wissensnutzung/-erzeugung und der Darstellung von Wissenstransfers unterschieden, da letztere nicht in allen Ansätzen explizit betrachtet werden, während die Ansätze grundsätzlich die Nutzung und Erzeugung von Wissen adressieren. Eine Ergänzung hierzu ist dann die Modellierung der Wissens-Ressourcen und ihrer Strukturen. Die Möglichkeit der Berücksichtigung von technischen Rahmenbedingungen (IuK-Technologien) und organisatorischen Rahmenbedingungen wird in den folgenden Tabellenzeilen betrachtet. Die letzte Zeile der Tabelle zielt auf den Umgang der Ansätze mit der Komplexität und Variabilität in den wissensintensiven Prozessen.

	KMDL	KPR	PROMOTE	DeCo	GPO-WM	KIPN
Zielstellung	Prozessanalyse und Verbesserung	Strategische Ausrichtung	Strategische Ausrichtung	Prozessspezifikation und -Analyse	WM-Einführung	Prozessspezifikation und -Analyse
Methodik	+	+	+	+	+	-
Werkzeugbindung	K-Modeler	ARIS	PROMOTE	Entwicklung ausstehend	-	-
Wissens - Nutzung/ Erzeugung	+	+	+	-	+	+
Wissens-Transfer	+	+	+	-	-	+
Wissens-Ressourcen	+	+	+	-	+	+
IuK-Technologie	-	+	-	-	+	-
Organisation	-	-	-	-	-	-
Komplexität und Variabilität	- Viele Details	o Keine Aussage	o Keine Aussage	+ Geringe Kontroll-flussorientierung	+ Hohe Abstraktion	+ Geringe Kontroll-flussorientierung

Tab. 2: Eigenschaften der Modellierungsansätze

4 Fazit

Gai & Dang weisen auf drei Limitierungen des prozessorientierten Wissensmanagements hin [GD10, S. 3-4]:

1. Nicht alle Wissensaktivitäten stehen mit Geschäftsprozessen in Verbindung. Ein Gegenbeispiel ist der Wunsch, während einer Teepause zu kommunizieren.
2. Die Veränderlichkeit der Prozesse wird durch viele Methoden nur unzureichend abgebildet. Wissensflüsse ändern sich und sind nicht an statische Prozesse gebunden.
3. Implizites Wissen wird oft unzureichend behandelt. Wissensträger werden nur als Attribut abgespeichert, dies reicht aber nicht aus, um den Wissensfluss darzustellen.

Punkt 1 trifft generell den Ansatz des geschäftsprozessorientierten Wissensmanagements. Der Kontext, insbesondere die organisatorischen Rahmenbedingungen haben einen erheblichen Einfluss auf das Funktionieren wissensintensiver Prozesse. Das trifft nicht nur die Durchführung von Wissensaktivitäten außerhalb der Prozesse. Die Modellierungsansätze greifen dies nicht auf (vgl. Tabelle 2). Der Kontext sollte jedoch bei der Modellierung berücksichtigt werden. Auch die Punkte 2 und 3 werden nur teilweise adressiert. Wie Tabelle 2 zeigt bilden nicht alle Modellierungsansätze explizit unterschiedliche Wege des Wissenstransfers ab. Für den Umgang mit der Komplexität und der Variabilität wissensintensiver Prozesse werden zwei grundsätzliche Wege gesucht - zum einen die Abkehr von der Kontrollflussorientierung und zum anderen eine hohe Abstraktion. Es zeigt sich, dass strategisch ausgerichtete Modellierungsansätze (GPO-WM, KPR) auf ein hohes Abstraktionsniveau setzen, während Ansätze zur detaillierten Spezifikation und Analyse von Prozessen (DeCo, KIPN) eine geringe Kontrollflussorientierung aufweisen. Neben dieser allgemeinen und offensichtlichen Unterscheidung bleibt die Frage, welcher Ansatz für welche Zielstellung am besten geeignet ist und wie sich die Modelle und Modellierungsaktivitäten in ein nachhaltiges Wissensmanagement einfügen. Wie sich Aufwand und Nutzen der Ansätze bewerten lassen.

Zusammenfassend lässt sich feststellen, dass mit Blick auf den Wissenstransfer im prozessorientierten Wissensmanagement lediglich Ideen und Hilfestellungen zu finden sind, aber kein vollständiger Ansatz. Häufig fehlen Betrachtungen für variable Prozesse, Ansätze zur Untersuchung der Organisation als solche und konkrete Vorgehensweisen. Weiterhin fehlen sowohl Betrachtungen des Aufwands als auch des erwarteten Nutzens gänzlich.

Literaturverzeichnis

[Al98] Allweyer, T.: Wissensmanagement mit ARIS-Modellen. In: Scheer, A.-W.: ARIS - vom Geschäftsprozeß zum Anwendungssystem, 3. Auflage, Springer Verlag, Berlin 1998, S. 162-168.

[BO14] BOC Group: Wissensmanagement mit PROMOTE, http://www.boc-group.com/de/landing-pages/wissensmanagement-mit-promoter/, (Datum des Zugriffs: 22.12.2014).

[FBS12] França, J.B.S., Baião, F.A., Santoro, F.M.: Towards Characterizing Knowledge Intensive Processes. In: Proceedings of the 2012 IEEE 16th International Conference on Computer Supported Cooperative Work in Design, 2012, S. 497-504.

[FBS13] França, J.B.S., Baião, F.A., Santoro, F.M.: A Notation for Knowledge-Intensive Processes. In: Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design, 2013, S. 190-195.

[GD10] Gai, Y., Dang, Y.: Process-oriented Knowledge Management: a Review on Strategy, Content, Systems and Processes. In: IEEE International Conference on Management and Service Science (MASS), August 2010.

[Gr09] Gronau, N.: Wissen prozessorientiert managen: Methode und Werkzeuge für die Nutzung des Wettbewerbsfaktors Wissen in Unternehmen, Oldenbourg Wissenschaftsverlag, München 2009.

[GMU04] Gronau, N., Müller, C., Uslar, M.: The KMDL Knowledge Management Approach: Integrating Knowledge Conversions and Business Process Modeling. In: Karagiannis, D., Reimer, U.: Practical Aspects of Knowledge Management. 5th International Conference, PAKM 2004, Wien Dezember 2004.

[He02] Heisig, P.: GPO-WM: Methode und Werkzeuge zum geschäftsprozessorientierten Wissensmanagement. In: Abecker et al.: Geschäftsprozessorientiertes Wissensmanagement, Springer Verlag, Berlin Heidelberg 2002, S. 47-64.

[HR07] Hepp, M., Roman, D.: An Ontology Framework for Semantic Business Process Management. In: Proceedings of Wirtschaftsinformatik 2007, Karlsruhe Februar - März 2007.

[HKT02] Hinkelmann, K., Karagiannis, D., Telesko, R.: PROMOTE: Methodologie und Werkzeug für geschäftsprozessorientiertes Wissensmanagement. In: Abecker et al.: Geschäftsprozessorientiertes Wissensmanagement, Springer Verlag, Berlin Heidelberg 2002, S. 65-90.

[Ka95] Karagiannis, D.: BPMS: Business Process Management Systems. In: ACM SIGOIS Bulletin, New York August 1995/Vol. 16, No. 1, S.10-13.

[KM14a] KMDL Blog, http://www.kmdl.de/ (Datum des Zugriffs: 24.11.2014).

[KM14b] K-Modeler | KMDL Blog, http://www.kmdl.de/?q=de/node/27 (Datum des Zugriffs: 24.11.2014).

[KBL13] Krallmann, H., Bobrik, A., Levina, O.: Systemanalyse im Unternehmen: Prozessorientierte Methoden der Wirtschaftsinformatik, 6. Auflage, Oldenbourg Verlag, München 2013.

[Le07] Lehner, F. et al.: Erfolgsbeurteilung des Wissensmanagements: Diagnose und Bewertung der Wissensmanagementaktivitäten auf Grundlage der Erfolgsfaktorenanalyse. In: Schriftenreihe Wirtschaftsinformatik, Diskussionsbeitrag W-24-07,3. Auflage, Passau 2007.

[MBG12] Müller, C., Bahrs, J., Gronau, N.: Considering the Knowledge Factor in Agile Software Development. In: Gronau, N. (Hrsg.): Modeling and Analyzing knowledge intensive business processes with KMDL - Comprehensive insights into theory and practice, Gito Verlag, Berlin 2012.

[NT95] Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company: How Japanese Companies create the Dynamics of Innovation, Oxford University Press, New York 1995.

[Po09] Pogorzelska, B.: Arbeitsbericht - KMDL® v2.2: Eine semiformale Beschreibungssprache zur Modellierung von Wissenskonversionen, 07. Januar 2009.

[Re02] Remus, U.: Prozeßorientiertes Wissensmanagement: Konzepte und Modellierung. Dissertation, Universität Regensburg, 2002.

[RN11] Rychkova, I., Nurcan, S.: Towards Adaptability and Control for Knowledge-Intensive Business Processes: Declarative Configurable Process Specifications. In: Proceedings of the 44th Hawaii International Conference on System Sciences, 2011.

[Sc98] Scheer, A.-W.: ARIS - vom Geschäftsprozeß zum Anwendungssystem, 3. Auflage, Springer Verlag, Berlin 1998.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Multi-Touch-Tisch oder Plastikfolie? – Design einer Studie zum Vergleich von Medien in der Modellierung

Anne Gutschmidt[1], Kurt Sandkuhl[2]

Abstract: Ein wichtiger Aspekt partizipativer Unternehmensmodellierung ist die Arbeit in der Gruppe. Doch verändert sich die Zusammenarbeit der Gruppenmitglieder in Abhängigkeit davon, welches Medium zur Erstellung der Modelle genutzt wird? Gibt es z.B. Unterschiede im Verhalten der Gruppe, wenn sie mit einer Plastikfolie, ähnlich einem Whiteboard, oder einem Multi-Touch-Tisch arbeitet? Ausgehend von Forschungsstand und theoretischen Grundlagen der Gruppenarbeit sowie bisherigen Forschungen zu Multi-Touch-Tischen werden in diesem Beitrag relevante Forschungsfragen aufgeworfen und ein experimentelles Design beschrieben, um mögliche medienabhängige Unterschiede der Gruppenarbeit zu untersuchen. Dabei spielen Aspekte wie Formen der Zusammenarbeit, z.B. verbale und nonverbale Beiträge der Beteiligten, aber auch Territorialverhalten und Gruppenleistung eine Rolle.

Keywords: Multi-Touch-Table, Plastikfolie, Unternehmensmodellierung, Medienvergleich

1 Einführung

Steht ein Unternehmen vor Veränderungen, die beispielsweise durch die Einführung neuer Produkte oder IT-Systeme, neue Konkurrenten oder auch neue gesetzliche Regelungen bedingt sein können, ist es wichtig, das Unternehmen in seiner Gesamtheit mit all seinen Abläufen, Strukturen und Abhängigkeiten zu kennen. Die Geschäftsprozess- und Unternehmensmodellierung stellen dabei wichtige Werkzeuge dar. Durch sie kann der Ist-Zustand einer Organisation visualisiert werden, so dass Ursachen für Probleme und Veränderungspotentiale häufig überhaupt erst sichtbar gemacht werden [SWS13]. Um einen maximalen Nutzen aus der Modellierung zu ziehen, ist es sinnvoll, die Betroffenen in den Modellierungsprozess einzubeziehen, d.h. Verantwortliche aus einzelnen Fachabteilungen und IT-Verantwortliche. Die partizipative Modellierung folgt genau diesem Ansatz. Sie soll dafür sorgen, dass die tatsächlich betroffenen Personen aktiv Modelle mitgestalten und die Modelle so im Ergebnis eine höhere Akzeptanz erreichen. Um die Fachexperten des Unternehmens bei der Modellierung zu unterstützen, werden außerdem Personen, die Expertise in der Modellierungssprache besitzen, als Moderatoren (sogenannte Facilitators) hinzugezogen, die die Diskussion leiten und Ergebnisse in digitaler Form dokumentieren [SPS07]. Damit ist Gruppenarbeit ein wesentlicher Aspekt der Prozess- und Unternehmensmodellierung.

[1] Universität Rostock, Institut für Informatik, Lehrstuhl für Softwaretechnik, Albert-Einstein-Straße 22, 18059 Rostock, anne.gutschmidt@uni-rostock.de
[2] Universität Rostock, Institut für Informatik, Lehrstuhl für Wirtschafsinformatik, Albert-Einstein-Straße 22, 18059 Rostock, kurt.sandkuhl@uni-rostock.de

Im Rahmen einer partizipativen Modellierungssitzung können ganz unterschiedliche Hilfsmittel eingesetzt werden. [SWS13] beschreiben den Einsatz von Plastikfolie als einen geeigneten Weg, um die Beteiligten zur aktiven Teilnahme zu animieren. Üblicherweise wird die Plastikfolie an eine Wand gehängt, wo sie dann, ähnlich wie bei einem Whiteboard, beschriftet und mit farbigen Karten versehen werden kann. Der Einsatz der Plastikfolie erfordert es jedoch, dass die Moderatoren nicht nur die Diskussion leiten, sondern auch darauf achten, dass die Notationsregeln der genutzten Modellierungssprache eingehalten und später das Ergebnis in eine digitale Form überführt wird. Daher wäre es wünschenswert über ein Modellierungswerkzeug zu verfügen, das sowohl die Einhaltung von Notationsregeln als auch die Ergebnisdokumentation erleichtert, ohne dass der partizipative Charakter in einer Modellierungssitzung verloren geht. Multi-Touch-Tische, auch Tabletops genannt, ermöglichen es, dass mehrere Nutzer gleichzeitig mit einer Software durch Berührung der Oberfläche interagieren [LR13]. Demnach könnten sie als alternatives Werkzeug gegenüber der Plastikfolie zur partizipativen Modellierung genutzt werden. Da es sich dabei um ein computergestütztes Werkzeug handelt, kann bei der Modellierung eine Entwicklungsumgebung bereitgestellt werden, die die Anwendung der Modellierungssprache vereinfacht. Darüber hinaus sind die Modelle bereits während der Sitzung digital gesichert.

Daraus ergibt sich die Frage, inwiefern sich ein Multi-Touch-Tisch im Gebrauch für die Modellierung von der traditionellen Plastikfolie unterscheidet. Eine empirische Studie soll klären, welche Unterschiede zwischen beiden Werkzeugen bei der partizipativen Modellierung bestehen. Hierbei soll der besondere Schwerpunkt darauf liegen, welche Unterschiede die Medien in der Gruppenarbeit hervorrufen und ob die konkrete Art und Weise, wie die Werkzeuge für die Gruppenarbeit eingesetzt werden, verschieden ist.

Gegenstand dieses Beitrags ist das Design eines Experiments, das den Vergleich von Multi-Touch-Tisch und Plastikfolie beim kollaborativen Modellieren in Gruppen erlaubt. Da Gruppenarbeit das zentrale Thema der Studie ist, werden im Abschnitt 2 der dazugehörige Forschungsstand sowie Theorien kurz beschrieben. Im Bereich der Nutzung von Multi-Touch-Tischen wurden bereits Studien durchgeführt. Abschnitt 3 wird dazu jedoch zeigen, dass diese Studien meist explorativer und noch häufiger deskriptiver Natur sind. Auch Studien mit experimentellem Design, die Multi-Touch-Tische mit Medien wie Papier oder PCs vergleichen, liefern noch kein zufriedenstellendes Bild von der Gruppenarbeit mit Tabletops. Sie bieten einerseits keine ausreichende Basis, um Schlussfolgerungen für den Vergleich mit einer Plastikfolie zu ziehen, und weisen andererseits teilweise widersprüchliche empirische Befunde auf. Im Anschluss an die Aufarbeitung von Theorie und Forschungsstand werden konkrete Forschungsfragen abgeleitet, die durch die Studie beantwortet werden sollen. In Abschnitt 4 wird dann das Forschungsdesign der Studie vorgestellt.

2 Gruppenarbeit

Die partizipative Modellierung ist eine klassische Gruppenarbeit, in der Fachexperten und Methodenexperten eine gemeinsame Aufgabe bearbeiten. In der Literatur werden unterschiedliche Eigenschaften aufgezählt, um Gruppen zu charakterisieren. Von einer Gruppe spricht man, sobald drei oder mehr Personen direkt und über eine längere Zeitspanne interagieren. Laut [Ner12] zeigen sich Gruppenphänomene tatsächlich erst ab einer Anzahl von drei Personen. Innerhalb einer Gruppe differenzieren sich über die Zeit Rollen heraus, wie z.B. die Rolle des Führers oder des Spezialisten. Damit die Personen wissen, wie sie sich in der Zusammenarbeit verhalten sollen, müssen zur Orientierung gemeinsame Normen, Werte und Ziele festgelegt werden. Die Gruppenmitglieder erleben und präsentieren sich typischerweise auch nach außen hin als zusammengehörig [vRN11; Sad91; Jon14].

Gruppenarbeit kann zu vielen Vorteilen führen, insbesondere wenn Aufgaben so komplex sind, dass sie einzeln nicht mehr bewältigt werden können. Allerdings wurde in vielen Fällen auch nachgewiesen, dass die Leistung einer Gruppe geringer sein kann, als wenn die Leistung durch die Mitglieder einzeln erbracht würde. Die tatsächliche Leistung weicht also von der potentiellen Leistung in der Weise ab, dass es zu Leistungsverlusten kommt, die Gruppenprozessen zugeschrieben werden. Dies ist u.a. abhängig von der Art der Aufgabe, die die Gruppe bearbeitet [Jon14]. Laut [Ste66] unterscheiden sich Aufgaben danach, ob sie sich unter den Mitgliedern aufteilen lassen, ob eher Quantität oder Qualität erzielt werden soll und in welcher Art und Weise die Gruppenmitglieder zum Gesamtergebnis beitragen. Gewinne und Verluste durch Gruppenprozesse entstehen im Bereich der Motivation, der individuellen Fertigkeit und der Koordination [DSSH11]. Social Facilitation und Social Compensation sind Beispiele für Prozessgewinne im Bereich der Motivation [vRN11]. Bei Social Facilitation führt die bloße Gegenwart anderer dazu, dass sich eine Person mehr anstrengt, als wenn sie allein wäre. Von Social Compensation spricht man, wenn ein stärkeres Gruppenmitglied versucht, die Leistungsdefizite schwächerer Gruppenmitglieder auszugleichen [vRN11; DSSH11]. Könnten Teilnehmer einer Modellierungssitzung also aus eventuell mangelnder Kenntnis weniger beitragen, würde sich ein Teilnehmer, der hier eine größere Expertise hat, noch mehr bemühen, als wenn er die Aufgabe alleine bearbeitet hätte, um Defizite der Kollegen auszugleichen.

Es kann aber auch zu Prozessverlusten bedingt durch eine geringere Motivation der Gruppenmitglieder kommen. Beim sogenannten Social Loafing leisten Gruppenmitglieder weniger als in Alleinarbeit, weil sie ihren eigenen konkreten Beitrag zur Gesamtlösung nicht erkennen können [KW93; vRN11; DSSH11].Wird jedoch darauf geachtet, dass die Einzelbeiträge sichtbar sind für die Gruppenmitglieder, hat dies wieder positiven Einfluss auf die Motivation. Dadurch soll es sogar zum sogenannten Köhler-Effekt kommen [KMPS05], der unter die Social Facilitation fällt und bei dem ein Gruppenmitglied mehr leistet, um nicht für eine schlechtere Gruppenleistung verantwortlich zu sein [DSSH11]. Sofern Farben in der Notation einer Modellierungssprache nicht bereits eine bestimmte Bedeutung zugeordnet ist, könnten Beiträge einzelner Gruppenmitglieder die

Farbe der genutzten Kärtchen oder Stifte identifizierbar bleiben.

Vom Free-Riding-Effekt oder auch Trittbrettfahren wird gesprochen, wenn Gruppenmitglieder absichtlich weniger leisten, weil sie glauben, kaum Einfluss auf das Ergebnis der Gruppenarbeit zu haben [Jon14; DSSH11]. Ein Teilnehmer einer Modellierungssitzung sollte also nicht das Gefühl haben, dass es egal ist, ob er an einer Lösung mitarbeitet oder nicht. Dies könnte z.B. in einer Modellierungssitzung geschehen, wenn das Modell am Computer mit nur einer Maus und Tastatur eingegeben würde, ähnlich wie in einer Studie von [RLHM09]. Die Autoren stellten fest, dass im Vergleich zur Nutzung eines Multi-Touch-Tischs Redebeiträge und physische Beiträge im Sinne von Eingaben unter den Gruppenmitgliedern, die nur einen Laptop nutzten, signifikant ungleichmäßiger verteilt waren. Die Beteiligung derjenigen, die weniger Zeit an den Eingabegeräten verbrachten, war geringer, da sie buchstäblich nur zuschauen konnten. Demgegenüber hatten diejenigen, die mehr Zeit an den Eingabegeräten hatten, auch mehr Redebeiträge, da sie den anderen offenbar ihre Aktionen erklärten.

Prozessverluste und -gewinne im Bereich der individuellen Fertigkeit können z.B. durch kognitive Einschränkung bzw. kognitive Stimulation entstehen [DSSH11; Jon14]. Im Bereich der Unternehmensmodellierung werden teilweise Ideen und Wissen im Stile des Brainstorming nach [Osb63] zusammengetragen. Einerseits können die Äußerungen eines Gruppenmitglieds die anderen in eine thematische Richtung lenken und damit ihr Denken einschränken. Andererseits können die Äußerungen die anderen auf Ideen bringen, die ihnen so unter Umständen nicht eingefallen wären [DSSH11].

Es kann auch zu Prozessverlusten durch Schwierigkeiten der Koordination kommen, z.B. wenn man beim Brainstorming ein Gruppenmitglied erst ausreden lassen muss, bevor man selbst seine Ideen äußern kann. Empirische Untersuchungen haben mehrfach belegt, dass beim Brainstorming tatsächlich weniger Ideen produziert werden, als wenn die Gruppenmitglieder ihre Ideen einzeln für sich festhalten [LT73] Prozessgewinne im Bereich der Koordination konnten bisher nicht nachgewiesen werden [DSSH11].

3 Multi-Touch-Table als Werkzeug in der Gruppenarbeit

Multi-Touch-Tische halten aufgrund der direkten und natürlichen Bedienweise und wegen ihrer sehr guten Visualisierungsmöglichkeiten in verschiedensten Bereichen Einzug, darunter vor allem als Unterrichtsmittel [PH09], aber auch in der Medizin [LRF+11] oder Architektur [MPWW07]. Die Größe des Mediums und Weiterentwicklungen der Dateneingabe [LR13] machen es vor allem möglich, dass mehrere Personen die Bedienoberfläche gleichzeitig benutzen können. Das macht sie für Aufgaben des kollaborativen Designs und Modellierens besonders geeignet. So finden sich Studien, die beschreiben, wie mit Hilfe eines Multi-Touch-Tisches öffentliche Plätze gestaltet werden [RLHM09; MPWW07] oder wie mit mehreren Personen Software entworfen wird, indem UML-Diagramme gezeichnet werden [BBB12a; PSTS12]. Auch für Aufgaben, in denen es allgemeiner um Kreativität und das Zusammentragen von Ideen geht, wie es im Brainst-

orming üblich ist, wurden Multi-Touch-Tische erprobt [SBC+12; BBAV12]

In diesem Abschnitt werden bisherige Erkenntnisse zu Territorialverhalten und Awareness (3.1), zum Verhalten in der Zusammenarbeit (3.2) und zur Leistung (3.3) vorgestellt.

3.1 Territorialverhalten und Awareness

[RFSM04] beschäftigten sich in einer Studie, in der Probanden Teile eines Gedichts wie ein Puzzle zusammensetzen sollten, mit der Größe von Tabletops. Sie stellten fest, dass selbst größere Gruppen in ihrer Studie nicht von einem größeren Tisch profitierten. Die Probanden mit dem größeren Tisch waren weder schneller, noch hatte die Tischgröße einen Einfluss auf die Arbeitsteilung oder die subjektive Bewertung durch die Probanden. Die Autoren erklären dies teilweise durch die Bildung von Territorien auf dem Tisch. [RFSM04] beobachteten, dass sich für Nutzer immer drei Arten von Territorien bildeten: ein Bereich, an den nur der jeweilige Nutzer herankommt, ein Bereich an den der Nutzer sowie alle anderen Nutzer herankommen und ein Bereich, an den nur die anderen Nutzer herankommen. [SC10] beschreiben ein ähnliches Territorialverhalten, bei dem es einen persönlichen Bereich gibt, auf den kein anderer Nutzer ungefragt zugreift, einen Bereich, auf den die gesamte Gruppe zugreift, sowie einen Bereich, auf den der jeweilige Nutzer und seine Nachbarn zugreifen. Letzterer dient eher dem Zwischenlagern von Arbeitsgegenständen. [RFSM04] und [SC10] schlussfolgerten gleichermaßen, dass, je weiter Objekte von einem Nutzer entfernt sind, umso weniger fühlt sich dieser für sie verantwortlich. Damit kann ein größerer Tisch dazu führen, dass auch der Bereich des Tisches größer wird, für den sich kein Nutzer verantwortlich fühlt.

Allerdings bietet ein Multi-Touch-Tisch, der groß genug ist, dass sich Nutzer private Bereiche einrichten, einige Vorteile. [DB92, S. 107] betonen die Wichtigkeit der Awareness, die sie als „an understanding of the activities of others, which provides a context for your own activity“ definieren. Zur Abstimmung in der Gruppenarbeit ist es demnach wesentlich, dass jeder vom anderen weiß, was dieser tut und warum. Ein privater Bereich auf dem Multi-Touch-Tisch ermöglicht Nutzern, Dinge außerhalb der Gruppe auszuprobieren, wobei die anderen ihnen immer noch über die Schulter schauen können. So bleiben die Gruppenmitglieder über die Aktivitäten informiert und können ihren Nachbarn gegebenenfalls Fragen stellen oder Hilfe leisten [SC10].

Eine Auslagerung auf externe Geräte wie Tablets oder andere mobile Geräte scheint naheliegend. Die Ergebnisse einer Studie von [BTKG11] deuten jedoch an, dass der Anreiz zur Zusammenarbeit am stärksten wirkt, wenn eine Gruppe ausschließlich an einem Tabletop arbeitet. [SC10] raten außerdem davon ab, fixe Bereiche auf einem Multi-Touch-Tisch einzurichten, da sich in ihren Beobachtungen die Territorien fortlaufend veränderten. Es gibt aber auch Design-Ansätze, die die zusätzliche Bereitstellung spezieller Features der Benutzeroberfläche vorsehen, um die Awareness sicherzustellen (siehe z.B. [IF09]). Plastikfolie ermöglicht im Gegensatz dazu nur eingeschränkte Möglichkeiten. Die Bildung und dynamische Veränderung von Territorien an der Plastikfolie selbst erscheint eher unwahrscheinlich, da Änderungen und Korrekturen aufwendiger sind.

[PH09] zeigten, dass Kinder zum Lösen von Aufgaben, Tabletops gegenüber Papier bevorzugten. Es lud mehr zum Ausprobieren ein, da sich Fehler leichter korrigieren ließen. Ähnliche Befunde zeigen sich in Studien von [BB12] bei dem Vergleich von Papier und Multi-Touch-Tisch für den Entwurf von UML-Diagrammen und in Interviews von [RLHM09] zum Entwurf einer Parkanlage u.a. mit Tabletops. Damit scheinen Tabletops durch ihre Dynamik einen Vorteil zu bieten. Zusammengefasst ist die Awareness ein wichtiger Aspekt in der Abstimmung unter den Gruppenmitgliedern und damit auch für die gesamte Zusammenarbeit, der beispielsweise durch die Möglichkeit des individuellen Ausprobierens in persönlichen, aber dennoch sichtbaren Bereichen unterstützt werden kann.

3.2 Verhalten in der Zusammenarbeit

In Studien, die sich speziell auf die Zusammenarbeit an Multi-Touch-Tischen widmen, ist für die Autoren vor allem von Interesse, wie eng die Zusammenarbeit der Gruppenmitglieder ist, ob die Personen parallel arbeiten und ob die Beiträge gleichmäßig verteilt sind. Diese Beiträge beziehen sich meist auf Redebeiträge, Gestik oder Interaktionen mit der Bedienoberfläche, um an der Aufgabe zu arbeiten. [BBAV12] zeigten in ihren Experimenten, dass Redebeiträge und Gestik am Tabletop in einer Brainstorming-Aufgabe geringer ausfielen als an einem mit Papier ausgelegten Tisch, obwohl die Autoren eine Software mit attraktiven Features auf dem Multi-Touch-Tisch zur Verfügung gestellt hatten. Flipcharts schnitten jedoch schlechter ab als Tabletops. [BBAV12] schlussfolgern aus ihrer Studie, dass die Ausrichtung der Arbeitsfläche einen wesentlichen Einfluss auf die Zusammenarbeit hat, aber dass die Attraktivität des Tabletops durch die bereitgestellten Features von der Zusammenarbeit eher abgelenkte. Auch [BB12] verglichen zunächst die Nutzung von Papier und Stift mit Multi-Touch-Tischen und kamen zu dem Schluss, dass die Probanden einen größeren Zeitanteil damit verbrachten, tatsächlich eng zusammenzuarbeiten. Im Gegensatz dazu arbeiteten die Gruppenmitglieder auf Papier mehr individuell für sich. [BBB12b] verglichen in einer weiteren Studie die Nutzung von PC und Multi-Touch-Tisch. Sie untersuchten, wie gleichmäßig die Beiträge in einem Zweierteam in Form von Eingaben bei der Erstellung von UML-Zustandsdiagrammen verteilt waren. Die Arbeitsbeiträge waren am Tabletop signifikant ausgeglichener. Darüber hinaus stellten sie fest, dass die Zusammenarbeit am Multi-Touch-Tisch enger war und am PC eher eine Person arbeitet, während die andere nur zusieht. Auch [RLHM09] kamen zu einem ähnlichen Ergebnis bei einem experimentellen Vergleich von Laptop, Tabletop und Tabletop in Verbindung mit Tafeln und Tangible Objects, also greifbaren Objekten, über die man mit dem Tabletop zusätzlich interagieren kann, wobei sich die Autoren neben Interaktionen mit der Bedienoberfläche auch auf Redebeiträge konzentrierten. Tabletops, insbesondere in Verbindung mit Tangible Objects, scheinen mehr zur Mitarbeit zu animieren und erleichtern es offenbar, auch andere Nutzer zum Mitmachen aufzufordern.

Die Autoren stellten fest, dass bei der greifbareren Schnittstelle Personen, die eher in geringerem Maßen durch Redebeiträge zur Arbeit beitrugen, dies mehr durch Zeigen und

Gestikulieren kompensierten. Dies könnte jedoch ein Vorteil der Plastikfolie sein, da die Personen hier mit Gegenständen wie farbigen Karten agieren und auf sich aufmerksam machen können. Insgesamt existieren vorrangig Studien, die Multi-Touch-Tische mit Papiermedien oder PCs vergleichen. Die Plastikfolie könnte sich jedoch als flexibler erweisen, da kleinere Korrekturen möglich sind, und das Medium durch den Einsatz zusätzlicher Hilfsmittel wie Karten greifbarer erscheint.

3.3 Leistung

Einige der oben erwähnten Studien, die Multi-Touch-Tische mit anderen Medien verglichen, untersuchten auch mögliche Unterschiede in der Gruppenleistung. Hinsichtlich der Leistung interessierte die Autoren meist, wie viel Zeit eine Gruppe für die Bearbeitung eines Problems benötigte, sowie die Qualität der Lösung. [Bas2013] berichtet, dass die Probanden bei der Erstellung von UML-Zustandsdiagrammen mehr Zeit benötigten, als sie den Tabletop nutzten, als wenn sie den PC nutzten. Der Autor führt dies vor allem auf die schwierigere Eingabe von Text mittels einer Touch-Tastatur zurück, da die Probanden mit der üblichen Tastatur am PC doppelt so schnell umgehen konnten. Auf der anderen Seite ist die Qualität der Lösungen signifikant besser am Multi-Touch-Tisch, gemessen durch zwei Softwareexperten auf einer Zehnerskala. Im Vergleich von Flipcharts und Tischen, die mit Papier ausgelegt wurden, mit Multi-Touch-Tischen für Brainstorming-Aufgaben stellten [BBAV12] keine Unterschiede in der Anzahl der in der Gruppe produzierten Ideen fest. Die Autoren bewerteten jedoch auch die von den Gruppen vorgenommenen Kategorisierungen der Ideen. Hier schnitt der Papiertisch am besten ab.

[IFM+10] unternahmen zwar keinen Vergleich von Multi-Touch-Tischen mit anderen Medien, stellten aber fest, dass die Güte einer Lösung bei einer Analyseaufgabe am Tabletop höher ist, wenn die Gruppe enger zusammenarbeitet.

4 Forschungsfragen und Methode

4.1 Forschungsfragen

Die vorangegangenen Abschnitte zeigen, dass empirische Untersuchungen zu Multi-Touch-Tischen meist beschreibender oder explorativer Natur waren. Experimentelle Studien in diesem Bereich lassen nur in geringem Maße Prognosen für den Vergleich von Plastikfolie mit Multi-Touch-Tischen zu. Nicht immer sind die den Probanden gestellten Aufgaben exakt vergleichbar mit Fragestellungen der Unternehmensmodellierung, auch wenn sich Elemente wie z.B. das Brainstorming durchaus wiederfinden lassen. Darüber hinaus wurden in den Studien häufig nur Gruppen bestehend aus zwei Personen untersucht, wie z.B. von [BBB12a], [BTKG11] und [IFM+10], wobei laut [Ner12] Gruppenphänomene erst ab einer Anzahl von drei Personen auftreten. Auch sind

die empirischen Befunde nicht immer eindeutig. Darüber hinaus ist die Plastikfolie nicht ganz vergleichbar mit Medien, die in den vorgestellten Studien Multi-Touch-Tischen gegenübergestellt wurden.

Da die bisherige Befundlage zum kollaborativen Modellieren an Multi-Touch-Tischen als noch nicht ausreichend zu erachten ist, wird eine explorative Untersuchung vorgeschlagen, mit deren Hilfe Hypothesen zu Unterschieden in der Nutzung von Multi-Touch-Tischen und Plastikfolie generiert werden. Um das Untersuchungsfeld näher abzugrenzen sollen jedoch zunächst konkrete Forschungsfragen herausgearbeitet werden. Wie bereits erwähnt wurde, ist das zentrale Thema der Studie die Zusammenarbeit. Hierbei ist von Interesse, wie gut und wie eng die Gruppenmitglieder zusammenarbeiten. Bisherige Untersuchungen betrachteten dazu, wie gleichmäßig die Beiträge zur Lösung einer Aufgabe auf die Personen verteilt waren. Um Hinweise auf eventuelle Prozessverluste, z.B. durch Koordinations- oder Motivationsverluste, zu erhalten, können aber auch subjektiv wahrgenommene Einzelbeiträge aufschlussreich sein. Besonders interessant dürfte jedoch sein, ob und wie das jeweilige Medium für die Zusammenarbeit eingesetzt wird. Gestikulieren Gruppenmitglieder z.B. mit Kärtchen oder bewegen sie Objekte auf dem Multi-Touch-Tisch, um Ideen zu verdeutlichen? Damit ergibt sich die erste Forschungsfrage mit den folgenden Teilfragen:

1. Gibt es einen Unterschied bei der partizipativen Unternehmensmodellierung zwischen Multi-Touch-Tisch und Plastikfolie in der Gruppenzusammenarbeit?

a) Gibt es Unterschiede in der tatsächlichen Aufteilung der Beiträge?

b) Gibt es Unterschiede in der von den Gruppenmitgliedern wahrgenommenen Aufteilung der Beiträge?

c) Gibt es Unterschiede in der Nutzung des Mediums für die Zusammenarbeit?

d) Gibt es Unterschiede in der Motivation der Gruppenmitglieder in Abhängigkeit vom Medium?

e) Gibt es Unterschiede in der Koordination der Gruppenmitglieder in Abhängigkeit vom Medium?

Für die Zusammenarbeit ist wichtig, dass Gruppenmitglieder darüber informiert sind, woran die anderen gerade arbeiten. In diesem Zusammenhang wird vom Begriff der Awareness gesprochen. Ob die Gruppenmitglieder sich informiert fühlen, können jedoch am besten sie selbst sagen, weshalb die subjektiv wahrgenommene Awareness von vorrangigem Interesse sein soll. Doch woran erkennen die Gruppenmitglieder, an was die anderen gerade arbeiten? Auch mit dieser Frage soll sich die Studie befassen. Die Bildung persönlicher Arbeitsbereiche, die trotzdem für andere sichtbar sind, kann hierbei möglicherweise helfen. Für zukünftige Untersuchungen könnte dies außerdem interessant sein, da hier motivationale Aspekte eine Rolle spielen könnten. Sieht ein Gruppenmitglied, was und vor allem dass andere auch etwas zur Lösung beitragen, so könnte der Effekt eintreten, dass dieses Gruppenmitglied nicht „hinterherhinken“ möchte, oder aber

es sieht dadurch die Schwächen der anderen und versucht diese im Sinne der Social Compensation zu unterstützen. Aktivitäten im eigenen Arbeitsbereich könnten aber auch zur Social Facilitation beitragen, d.h., die eigenen Leistungen steigern sich, weil sie für andere sichtbar sind. Dies kann mit der ersten Studie sicherlich nicht im Detail geklärt werden, aber es könnten sich Hinweise finden lassen. Deshalb sollen sowohl Awareness als auch die Bildung von Territorien explorativ mit den folgenden Forschungsfragen untersucht werden:

2. Gibt es Unterschiede bei der partizipativen Unternehmensmodellierung zwischen Multi-Touch-Tisch und Plastikfolie in der Bildung von Territorien?

3. Gibt es Unterschiede bei der partizipativen Unternehmensmodellierung zwischen Multi-Touch-Tisch und Plastikfolie in der wahrgenommenen Awareness der Personen bezüglich der anderen Gruppenmitglieder?

4. An welchen Hinweisen orientieren sich die Personen bei Plastikfolie und Multi-Touch-Tisch, um über ihre Kollegen informiert zu bleiben?

Wie erwähnt, wurden Multi-Touch-Tische bisher mit Medien wie PCs, Papiertischen und Flipcharts verglichen. Papiertische und Flipcharts sind der Plastikfolie nur in Teilen ähnlich. Wie bei Flipcharts können zusätzlich Karten genutzt werden, und auch die Ausrichtung der Arbeitsfläche ist ähnlich, aber auf der Plastikfolie sind Korrekturen möglich – wenn vielleicht auch nicht ganz so leicht und schnell wie am Multi-Touch-Tisch. Daher kann nicht genau vorhergesagt werden, wie sich Multi-Touch-Tisch und Plastikfolie hinsichtlich der Leistung, im Sinne von Geschwindigkeit und Qualität unterscheiden werden. Hieraus ergeben sich die letzten Forschungsfragen:

5. Gibt es Unterschiede bei der partizipativen Unternehmensmodellierung zwischen Multi-Touch-Tisch und Plastikfolie hinsichtlich der Gruppenleistung?

a) Gibt es Unterschiede in der Aufgabendauer?

b) Gibt es Unterschiede in der Qualität der Lösung?

4.2 Methode

Wie im vorherigen Abschnitt bereits erwähnt wurde, wird für den empirischen Vergleich von Multi-Touch-Tisch und Plastikfolie eine explorative Studie vorgeschlagen. Das bedeutet, dass im Vornherein keine Hypothesen aufgestellt werden, da bisherige empirische Befunde und Theorien nicht genug Anhaltspunkte liefern. Vielmehr ist es der Zweck der Studie, das Untersuchungsfeld zu erkunden und, wenn möglich, Hypothesen aufzustellen, die in zukünftigen explanativen Studien geprüft werden müssen [BD06].

Stichprobe und Ablauf

Für die Studie wird eine Gruppengröße von drei Personen gewählt, da ab dieser Anzahl

Gruppenphänomene auftreten [Ner12]. Tatsächlich sollen in der Praxis mehrere Betroffene in die Unternehmensmodellierung eingebunden werden [SWS13]. Die Größe des für die Studie zur Verfügung stehenden Multi-Touch-Tischs ist jedoch ein limitierender Faktor, da alle Personen noch bequem daran Platz finden sollen.

Als Probanden sollen Studenten der Wirtschaftsinformatik und Informatik rekrutiert werden, die mit der Modellierungssprache 4EM vertraut sind. Ziel ist es zunächst, eine homogene Stichprobe zu schaffen, um mögliche Einflüsse, die durch Stichprobeneigenschaften entstehen könnten, minimal zu halten. Die Probanden sollen jedoch zumindest mit einem Teil der Zielgruppe übereinstimmen. Das Design soll insgesamt einem Experiment entsprechen. Dabei wird eine unabhängige Variable manipuliert, so dass die Auswirkungen auf eine oder mehrere abhängige Variablen gemessen werden können. Das experimentelle Design ist die einzige Möglichkeit, kausale Zusammenhänge nachzuweisen [BD06]. Damit also die Auswirkung des Modellierungswerkzeugs auf abhängige Variablen wie die Leistung oder die Zusammenarbeit untersucht werden kann, wird zunächst das Modellierungswerkzeug manipuliert: Modellierungsaufgaben werden entweder mit der Plastikfolie bearbeitet oder mit dem Multi-Touch-Tisch. Nach der randomisierten Auswahl des Arbeitswerkzeugs wird der Gruppe eine Aufgabe aus dem Bereich der Ziel- und Problemmodellierung gestellt. Um die Probanden zu motivieren, wird angekündigt, dass die Ergebnisse später offiziell präsentiert werden müssen. Es wird ein Messwiederholungsdesign vorgeschlagen, bei dem die Gruppen zu zwei verschiedenen Terminen jeweils eine Aufgabe bearbeiten sollen. Bearbeitet dieselbe Gruppe zwei Aufgaben, ist es wahrscheinlich, dass sich Lerneffekte einstellen. Um diesen Einfluss zu minimieren, wird die Reihenfolge der eingesetzten Medien für jede Gruppe randomisiert, d.h., es wird zufällig entschieden, ob eine Gruppe zuerst die Plastikfolie und dann den Tabletop nutzt oder umgekehrt.

Für die Messung der abhängigen Variablen ist es notwendig, mehrere Erhebungsinstrumente zu benutzen. Dabei handelt es sich einerseits um die Verhaltensbeobachtung und andererseits um die Befragung sowohl durch halbstrukturierte Interviews als auch durch Fragebögen. Tabelle 1 zeigt zusammenfassend alle abhängigen Variablen, entsprechend den Forschungsfragen angeordnet, und, wie diese erhoben werden sollen.

Beobachtung als Erhebungsmethode

Von den geplanten Erhebungsmethoden soll an dieser Stelle nur auf die Beobachtung eingegangen werden, da sie als am wichtigsten erscheint, um vor allem die Zusammenarbeit in der Gruppe und mit Plastikfolie und Multi-Touch-Tisch zu erkunden. In ähnlichen Studien (z.B. [BBB12a]; [IFM+10]; [BBAV12]) beobachteten die Autoren ihre Probanden mit Videokameras, um Details der Zusammenarbeit zu untersuchen. In der Regel wird mehr als ein Beobachter eingesetzt, um die Reliabilität, d.h. die Messgenauigkeit, zu erhöhen [Ner12; BD06]. Mithilfe der Videoaufzeichnungen kann das Verfahren wesentlich vereinfacht werden, da die Beobachter das Material unabhängig voneinander und beliebig häufig sichten können.

Tabelle 1: Übersicht der abhängigen Variablen und wie sie gemessen werden, geordnet nach den zugehörigen Forschungsfragen.

Abhängige Variable	**Erhebungsmethode**
1. a) Beiträge (Arten, Aufteilung etc.)	Beobachtung
1. b) wahrgenommener Beitrag	Fragebogen
1. c) Mediennutzung für Beiträge	Beobachtung
1. d) Motivation	Fragebogen
1. e) Koordination	Fragebogen, Interview
2. Territorialverhalten	Beobachtung
3. Awareness	Fragebogen, Interview
4. Awareness-Hinweise	Interview
5. a) Aufgabendauer	Zeitmessung
5. b) Qualität der Lösung	Bewertung anhand Expertenkriterien

Wissenschaftliche, strukturierte Beobachtungen werden anhand eines Beobachtungsschemas durchgeführt [Ner12]; BD06]. Während [BBB12a] und [IFM+10] das beobachtete Verhalten nach einem Schema von [TTP+06] kodierten, welches wiederum induktiv von den Forschern gebildet wurde, empfiehlt es sich, ein etablierteres Beobachtungsschema für Gruppenarbeit anzuwenden. Hierzu bietet sich beispielsweise die Interaktionsprozessanalyse von [Bal50] an, bei der Handlungen nach 12 Kategorien, wie z.B. „Stimmt zu“ und „Fragt nach Meinung“ unterschieden werden. Letztendlich kann mit einem Beobachtungsschema auf Gruppenebene oder Personenebene ausgezählt werden, wie oft ein bestimmtes Verhalten auftrat.

In Bezug auf die einzelnen Handlungen der Beteiligten ist von Interesse, ob diese konkrete Beiträge zur Lösung sind, ob diese verbaler oder nonverbaler Form sind und von wem sie stammen. Auf diese Weise kann analysiert werden, wie ausgewogen die Beiträge unter den Mitgliedern geleistet werden. Darüber hinaus kann auf dieser Basis analysiert werden, inwiefern das jeweilige Medium bei diesen Handlungen genutzt wird. So können konkrete Nutzungsweisen ermittelt werden, die vielleicht von der Plastikfolie auf den Tabletop übertragen werden können, oder es zeigen sich ganz neue Nutzungsformen am Tabletop.

Zur Erfassung der möglichen Bildung von Territorien auf der Arbeitsfläche soll ein Vorgehen in Anlehnung an [SC10] gewählt werden. Bei der Beobachtung wird die Arbeitsfläche durch ein (gedachtes) Gitter in mehrere Felder eingeteilt. Anhand des Videomaterials wird analysiert, in welchen Feldern die einzelnen Probanden wie oft – gemessen an der Häufigkeit der Interaktionen – arbeiten. So sollte sich zeigen lassen, ob persönliche Bereiche gebildet werden, in denen vorrangig ein bestimmter Proband arbeitet. Zur Auswertung des Videomaterials werden demnach mehrere Personen benötigt, die zuerst einer intensiven Beobachterschulung unterzogen werden müssen.

4.3 Laborumgebung

Für die Durchführung des Experiments steht an der Universität Rostock eine Laborumgebung bereit, die nicht nur unterschiedliche Modellierungsinstrumente sondern auch das Aufzeichnen und Auswerten der Aktivitäten im Labor erlaubt. Die Grundidee des Labors ist es, die Durchführung aller Modellierungsphasen in Teamgrößen bis zu 6 Personen durchführen und hinsichtlich der Einzelaktivitäten dokumentieren und analysieren zu können. Es soll in Nutzerstudien dazu verwendet werden, Verbesserungsbedarfe in den einzelnen Modellierungsphasen aufzudecken, Einflüsse bestimmter Methoden und Technologien auszuwerten und innovative Ansätze zu realisieren und evaluieren zu können. Zu diesem Zweck müssen sowohl industriell akzeptierte Werkzeuge vorhanden sein, um praxisnah arbeiten zu können, als auch neue Technologien mit Innovationspotential für die Modellierung untersucht werden. Von besonderem Interesse ist der Übergang zwischen den Modellierungsphasen, insbesondere wenn dort Medienbrüche auftreten oder Paradigmenwechsel stattfinden.

Die folgende Tabelle stellt die Ausstattung der integrierten Modellierungsumgebung im Hinblick auf die unterschiedlichen Modellierungsphasen dar. Dies wird hinsichtlich der industriell etablierten Werkzeuge, innovativer Technologien und Werkzeuge für die Dokumentation und Analyse für die einzelnen Phasen aufgeschlüsselt, wobei verschiedene Technologien für mehrere Phasen geeignet sind.

Die Ausstattung erlaubt das konventionelle Modellieren (Plastikfolie, Whiteboard, Smartboard), das mobile Modellieren mit iPads, Modellieren am MultiTouch 55“ TouchTable und die Nutzung neuer Interaktionsgeräte (Gestensteuerung, Datenbrillen). Das Labor ist mit Kameras, Mikrofonen, Aufzeichnungsgeräten und Software für die Analyse und das Kodieren des aufgezeichneten Materials ausgestattet. Außerdem stehen Modellierungswerkzeuge aus der industriellen Praxis zur Verfügung, wie Troux Architect, ARIS, 4EM, planningIT oder Valuemation.

5 Ausblick

Langfristiges Ziel soll die Entwicklung einer Theorie zum kollaborativen Modellieren mit Tabletops verfolgt werden. Dies macht es notwendig, dass mehrere Erhebungen durchgeführt werden, in denen schrittweise mögliche Einflussgrößen systematisch beeinflusst werden, ähnlich wie beim Vorgehen der Grounded Theory in Verbindung mit dem Theoretical Sampling [BD06]. Interessant dürfte z.B. die Erprobung von Ausgestaltungsmöglichkeiten der Sichtbarkeit der Beiträge sein und deren Einfluss auf die Motivation oder auch eine Weiterverfolgung des Themas Awareness.

Tabelle 2: Ausstattung der Laborumgebung und unterstützte Modellierungsphasen

Ausstattung / *Phase*	**Industriell etablierte Werkzeuge**	**Technologien mit Innovationspotential**	**Werkzeuge zu Dokumentation und Analyse**
Abgrenzung & Zieldefinition	Whiteboard, Moderatorenkoffer	SmartBoard, Mobiles Modellieren	Videoaufzeichnung, Bilddokumentation
Initiale fachlich / konzeptuelle Modellierung	Whiteboard, Moderatorenkoffer	SmartBoard Tabletop	Videoaufzeichnung, Tracking Software
Fachlich / konzeptionelle Modellierung	Modellierungs-Tool (TopBraid, Troux)	SmartBoard Tabletop Mobiles Modellieren	Videoaufzeichnung, Software zum Test der Benutzerfreundlichkeit
Modellverifikation und -korrektur	Whiteboard + Papier; Modellierungs-Tool + Beamer oder SmartBoard	Digital Pen Touch Table Mobiles Modellieren	Videoaufzeichnung, Software zum Test der Benutzerfreundlichkeit
Entwicklung operativer Modelle	Workflow-Tool (Troux, ARIS)		Software zum Test der Benutzerfreundlichkeit
Wartung und Evolution von Modellen	Workflow Tool + Software Entwicklungsumgebungen	Digital Pen Mobiles Modellieren	Software zum Test der Benutzerfreundlichkeit

Danksagung

Die Laborumgebung für die geplante Studie wurde im Rahmen des Projekts „Integrierte Umgebung für die Unternehmens- und Wissensmodellierung“ durch das Land Mecklenburg-Vorpommern mit Mitteln des Europäischen Fonds für regionale Entwicklung (EFRE) unterstützt.

Literatur

[Bal50] Robert F. Bales: Interaction Process Analysis. A Method for the Study of a small Group. Chicago: The University of Chicago Press., 1950.

[Bas13] Basheri, M.: Multi-Touch Table for Enhancing Collaboration during Software Design. PhD thesis, Durham University, 2013.

[BB12] Basheri, M.; Burd, L.: Exploring the significance of multi-touch tables in enhancing collaborative software design using UML. In Frontiers in Education Conference (FIE), 2012, 1–5, 2012.

[BBAV12] Buisine, S.; Besacier,G.; Aoussat, A.; Vernier, F.:. How do interactive tabletop systems influence collaboration? Computers in Human Behavior, 28(1):49 – 59, 2012.

[BBB12a] Basheri, M.; Burd, L.; Baghaei, N.: Collaborative software design using multi-touch tables. In Engineering Education (ICEED), 1–5, Dec 2012.

[BBB12b] Basheri,M.; Burd, L.; Baghaei, N.: A multi-touch interface for enhancing collaborative uml diagramming. In Proceedings of the 24th Australian Computer-Human Interaction Conference, OzCHI '12, 30–33, New York, NY, USA, 2012. ACM.

[BD06] Bortz, J.; Döring, N.: Forschungsmethoden und Evaluation - für Human- und Sozialwissenschaftler. Springer Berlin Heidelberg,Wiesbaden, 4. überarb. Aufl., 2006.

[BTKG11] Bachl, S.; Tomitsch, M.; Kappel, K.; Grechenig, T.: The effects of personal displays and transfer techniques on collaboration strategies in multi-touch based multi-display environments. In Human-Computer Interaction–INTERACT 2011, 373–390. Springer, 2011.

[DB92] Dourish, P.; Bellotti, V.: Awareness and coordination in shared workspaces. In Proceedings of the 1992 ACM Conference on Computer- supported Cooperative Work, CSCW '92, pages 107–114, New York, NY, USA, 1992. ACM.

[DSSH11] Drewes, S.; Schultze, S.; Schulz-Hardt, S.: Sozialpsychologie – Interaktion und Gruppe, Kap. Leistung in Gruppen, 221–244. Hogrefe Verlag, Göttingen, 2011.

[IF09] Isenberg, P.; Fisher, D.: Collaborative brushing and linking for co-located visual analytics of document collections. Computer Graphics Forum, 28(3):1031–1038, 2009.

[IFM+10] Isenberg, P.; Fisher, D.; Morris, M.R.; Inkpen, K.; Czerwinski, M.: An exploratory study of co-located collaborative visual analytics around a tabletop display. In Visual Analytics Science and Technology (VAST), 179–186, 2010.

[Jon14] Jonas, K.W.S.M.H.: Sozialpsychologie. Springer-Lehrbuch. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[KMPS05] Kerr, N. L.; Messé, L. A.; Park, E. S.; Sambolec, E. J.: Identifiability, performance feedback and the Köhler effect. Group Processes & Intergroup Relations, 8(4):375–390, 2005.

[KW93] Karau, S.J.; Williams, K.D.: Social loafing: A meta-analytic review and theoretical integration. Journal of Personality and Social Psychology, 65(4):681–706, 1993.

[LR13] Laufs, U.; Ruff, C.: Aufbau von Multi-Touch-Systemen. In: Schlegel, T. (Hrsg.): Multi-Touch, Xpert.press, 145–152. Springer, Berlin Heidelberg, 2013.

[LRF+11] Lundstrom, C.; Rydell, T. ; Forsell, C.; Persson, A.; Ynnerman, A.: Multitouch table system for medical visualization: Application to orthopedic surgery planning. Visualization and Computer Graphics, IEEE Transactions on, 17(12):1775–1784, Dec 2011.

[LT73] Lamm, H.; Trommsdorff, G.: Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review. European Journal of Social Psychol-

ogy, 3(4):361–388, 1973.

[MPWW07] Maquil, V.; Psik, T.; Wagner, I.; Wagner, M.: Expressive interactions - supporting collaboration in urban design. In Proceedings of the 2007 International ACM Conference on Supporting Group Work, GROUP ’07, pages 69–78, New York, NY, USA, 2007. ACM.

[Ner12] Nerdinger, F.W.: Grundlagen des Verhaltens in Organisationen. Organisation und Führung. Kohlhammer, 2012.

[Osb63] Osborn, A.F.: Applied imagination; principles and procedures of creative problem-solving -. Scribner, New York, 1963.

[PH09] Piper, A.M.; Hollan, J.D.: Tabletop displays for small group study: Affordances of paper and digital materials. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’09, 1227–1236, New York, NY, USA, 2009. ACM.

[RFSM04] Ryall, K.; Forlines, C.; Shen, C.; Ringel Morris, M.: Exploring the effects of group size and table size on interactions with tabletop shared-display groupware. In Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, CSCW ’04, 284–293, New York, NY, USA, 2004. ACM.

[RLHM09] Rogers, Y.; Lim,Y.; Hazlewood, W.R.; Marshall, P.: Equal opportunities: Do shareable interfaces promote more group participation than single user displays? Human Computer Interaction, 24(1-2):79–116, 2009.

[Sad91] Sader, M.: Psychologie der Gruppe. Grundlagentexte Psychologie. Juventa-Verlag, 1991.

[SBC+12] Schmitt, L.; Buisine, S.; Chaboissier, J.; Aoussat, A.; Vernier, F.: Dynamic tabletop interfaces for increasing creativity. Computers in Human Behavior, 28(5):1892 – 1901, 2012.

[SC10] Scott, S.D.; Carpendale, S.: Theory of tabletop territoriality. In Müller-Tomfelde,C. (Hrsg.), Tabletops - Horizontal Interactive Displays, Human-Computer Interaction Series, 357–385. Springer, London, 2010.

[SPS07] Stirna, J.; Persson, A.; Sandkuhl, K.: Participative enterprise modeling: Experiences and recommendations. In Advanced Information Systems Engineering, 19th International Conference, CAiSE 2007, Trondheim, Norway, June 11-15, 2007, Proceedings, 546–560, 2007.

[Ste66] Steiner, I.D.: Models for inferring relationships between group size and potential group productivity. Behavioral Science, 11(4):273–283, 1966.

[SWS13] Sandkuhl, K.; Wißotzki, M.; Stirna, J.: Unternehmensmodellierung: Grundlagen, Methode und Praktiken. Xpert.press. Springer Berlin Heidelberg, 2013.

[TTP+06] Tang, A.; Tory, M.; Po, B.; Neumann, P.; Carpendale, S.: Collaborative coupling over tabletop displays. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’06, 1181–1190, New York, NY, USA, 2006. ACM.

[vRN11] von Rosenstiel, L.; Nerdinger, F.W.: Grundlagen der Organisationspsychologie: Basiswissen und Anwendungshinweise. Schäffer-Poeschel, 2011.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Wie erleben Anwender ihre Geschäftsprozesse? User Feedback mittels Mobile App

Michael Gebhart[1], Marco Mevius[2], Peter Wiedmann[3]

Abstract: Durch den Einsatz von mobilen Endgeräten (z.B. Tablets, Smartphones) erschließen sich immer mehr Möglichkeiten, die Ausführung von Geschäftsprozessen zu unterstützen. Beispielsweise können Geschäftsprozessaktivitäten (z.B. Genehmigung eines Angebots) ortsunabhängig bearbeitet werden, wodurch die Durchlaufzeit signifikant reduziert wird. Die Nutzung von mobilen Apps beschränkt sich hierbei meist nur auf die Unterstützung von effizienter und flexibler Interaktion zwischen den verschiedenen ausführenden Rollen. Dieser Artikel beschreibt, wie mobile Apps nicht nur die Ausführung, sondern auch die Optimierung von Geschäftsprozessen unterstützen können. Hierzu werden vordefinierte Qualitätskriterien kontextabhängig während der Ausführung von Aktivitäten erfasst. Die durch traditionelle Methoden erfassten Daten (z.B. Messung von Kennzahlen) werden somit durch in Echtzeit gesammeltes User Feedback ergänzt. Der Ansatz wird am Beispiel einer eigens entwickelten mobilen App demonstriert und evaluiert.

Keywords: Geschäftsprozessmanagement, Modellqualität, BPMN, Mobile App, User Feedback

1 Einleitung

Mobile Endgeräte (z.B. Tablets, Smartphones) sind heutzutage fester Bestandteil des täglichen (Arbeits)-Lebens [Te13] [In14]. Ständige Erreichbarkeit, ortsunabhängige Interaktion und schnelle Reaktionsfähigkeit sind Beispiele für den Einfluss der mobilen, digitalen „Begleiter“. Insbesondere bei flexiblen und zeitkritischen Geschäftsprozessen können mobile Endgeräte eine signifikante Effizienzsteigerung hervorrufen. Beispielsweise können Außendienstmitarbeiter direkt über neue Aktivitäten mobil informiert werden, ohne dass es zu zeitlichen Verzögerungen und damit zu Kundenunzufriedenheit führt. Ein Geschäftsprozess beschreibt eine Menge von manuellen, teilautomatisierten oder automatisierten Aktivitäten, welche durch (menschliche) Ressourcen ausgeführt werden (vgl. [Ob96] [Me06]). Als Aufgabe des Geschäftsprozessmanagements wird die Analyse, Modellierung, Implementierung, Ausführung und Optimierung von Geschäftsprozessen in der Literatur genannt (vgl. [ATW03] [AM11] [We12] [MOW14]).

Vom Brocke & Rosemann [BR10] beschreiben Informationstechnologie (IT) – und

[1] iteratec GmbH, Zettachring 10, 70567 Stuttgart, michael.gebhart@iteratec.de
[2] HTWG Konstanz, kips, 78462 Konstanz, mmevius@htwg-konstanz.de
[3] AXON IVY AG, Elsenheimerstraße 57, 80687 München, peter.wiedmann@axonivy.com

damit verbunden mobile Applikationen – als ein essentielles Element des Geschäftsprozessmanagements. Insbesondere zur Modellierung und Unterstützung der Ausführung von Geschäftsprozessen beschreiben existierende Ansätze des Geschäftsprozessmanagements bereits den Einsatz mobiler Endgeräte. Beispielsweise können mobile Apps wie das „BPM Touch“ [ba14] zur mobilen Erfassung und Modellierung von Geschäftsprozessen eingesetzt werden. Als logische Fortsetzung der mobilen Unterstützung des Geschäftsprozessmanagements beschreibt dieser Artikel die Optimierung von Geschäftsprozessen mittels mobilen Apps. Hierfür werden User Feedback in Echtzeit erfasst und hierdurch Qualitätskriterien ermittelt, die über bislang aus Modellen und Laufzeitinformationen extrahierten Kriterien hinausgehen. Beispielsweise werden Informationen über die ausführender User erfasst (z.B.“verursacht die Ausführung der Aktivität Stress?“). Unter der Annahme (vgl. [MSW13] [MOW14] [BR10]), dass für das Geschäftsprozessmanagement die gleichen Anforderungen und Ziele gelten, wie in der Informationstechnologie (z.B. Optimierung der Durchlaufzeit, Senkung der Kosten,…), werden Qualitätsmerkmale aus der Softwareentwicklung, die mittels User Feedback erfasst werden können (und aus dem Bereich User Experience (UX)), analysiert und adaptiert. Der Artikel beschreibt, wie Feedback während der Ausführung von Geschäftsprozessen erfasst werden kann. Hierdurch wird die bisherige Überwachung von Geschäftsprozessen um die Erfassung von Feedback und den damit gemessen weichen und harten Qualitätsmerkmalen der User Experience ergänzt. Die Messung erweitert die statisch aus den Geschäftsprozessmodellen ermittelbaren „harten“ Indikatoren (z.B. Durchlaufzeit). Beispielsweise wird ermittelt, wie zufrieden der User mit der Ausführung der Geschäftsprozessaktivität ist, wie er sich fühlt, wie ungestört er arbeiten kann etc. Um die agile (z.B. ortsunabhängige) Erfassung zu gewährleisten, wird eine mobile App implementiert und bei der Ausführung eingesetzt.

Der Artikel ist wie folgt gegliedert: In Abschnitt 2 werden verwandte Ansätzen aus der Literatur vorgestellt und bewertet. Abschnitt 3 beschreibt die Architektur und Implementierung der mobilen App. In Abschnitt 4 wird ein Anwendungsszenario skizziert und die Evaluationsergebnisse bewertet. Zum Abschluss wird in Abschnitt 5 der Artikel zusammengefasst und ein Ausblick auf weitere Forschungsarbeiten gegeben.

2 Literaturreview

Dieser Abschnitt dient zur Einordnung und Abgrenzung des in diesem Artikel beschriebenen Ansatzes. Hierzu werden sowohl bestehende Ansätze des Geschäftsprozessmanagements, als auch existierende Qualitätsmodelle und Qualitätsfragen beschrieben und bewertet.

2.1 Geschäftsprozessmanagement

Die Abbildung 1 stellt den Zyklus des Geschäftsprozessmanagements dar. Die in der Abbildung 1 visualisierten Phasen (Analyse/Optimierung, Modellierung/Implementierung, Ausführung/Feedback) orientieren sich an dem Lebenszyklus von Geschäftsprozessen (vgl. [ATW03] [AM11] [We12] [MOW14]). Beispielsweise wird ein Modell eines Geschäftsprozesses mit Hilfe einer grafischen Modellierungssprache (z.B. BPMN 2.0 [OM11]) modelliert. Auf Basis des Modells werden die Aktivitäten des Geschäftsprozess manuell, teilautomatisiert oder automatisiert ausgeführt. Daten (z.B. Messwerte von Kennzahlen, Dokumente,…), welche durch die Ausführung generiert werden, bilden das Fundament für eine Analyse und Optimierung des Geschäftsprozesses. In den einzelnen Phasen bestehen unterschiedliche Möglichkeiten, User Feedback zu erfassen, wobei dieser Artikel auf die Erfassung von User Feedback während der Ausführung von Aktivitäten fokussiert ist (vgl. Abbildung 1).

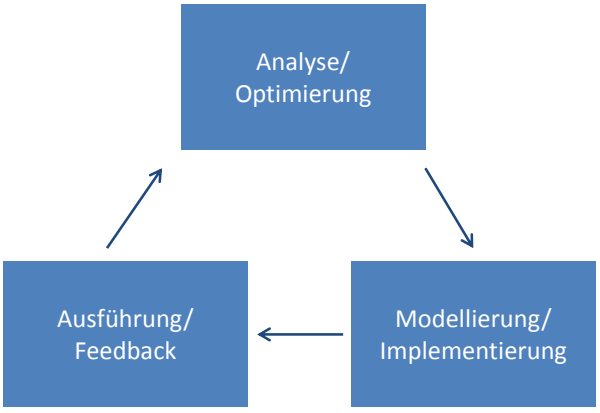


Abb. 1: Geschäftsprozessmanagementzyklus (eigene Darstellung)

Für die Phase der Modellierung und Implementierung existieren verschiedene Ansätze, User Feedback zu erfassen und zu nutzen.

Beispielsweise beschreiben Koschmider & Oberweis ein System, welches basierend auf dem Feedback eines Users die Absichten des Users analysiert. Als Ergebnis der Analyse können User, die von dem System vorgeschlagenen Modellierungsmöglichkeiten einsetzen. Das System ermöglicht somit eine schnelle und einfache Modellierung von Geschäftsprozessen [KO15].

Ein weiterer Ansatz wird von Mevius et al. [MOW15] vorgestellt. Durch eine agile Vorgehensweise werden in kurzen Abständen mit Hilfe von „Easy Capture Sheets“ alle am Geschäftsprozess beteiligten Rollen (Personen) befragt. Dieses Feedback wird in der nächsten Iteration des Geschäftsprozessmanagementzyklus analysiert und zur möglichen Optimierung verwendet. Mendling et al. [ULM15] führen ein Geschäftsprozessmodell-Matching auf Basis von User Feedback ein. Hierbei können beispielsweise zwei existierende Geschäftsprozessmodelle miteinander verglichen werden. Das Matching erfolgt in einem iterativen Prozess. Ein Matching-Algorithmus schlägt ein mögliches Matching vor, welches im zweiten Schritt durch ein User Feedback bestätigt oder verworfen wird. Durch die Optimierung des Matchings können zum Beispiel „doppelte“ Aufwände bei der Modellierung vermieden werden. In den Ansätzen von [KO15], [MOW15] und [ULM15] wird User Feedback zur Unterstützung der Modellierung von Geschäftsprozessen genutzt, sodass die anderen Phasen des Geschäftsprozessmanagements nicht berücksichtigt werden.

Jiménez-Ramírez et al. [Ji14] beschreiben einen Ansatz, welcher für deklarative Geschäftsprozessmodelle während der Ausführung eingesetzt werden kann. Deklarative Geschäftsprozessmodelle dokumentieren, im Gegensatz zu imperativen Geschäftsprozessmodellen, “WAS” (Ziel) und nicht “WIE” (in welcher Reihenfolge) etwas ausgeführt werden soll. Aufgrund der hohen Anzahl von Ausführungsvarianten deklarativer Geschäftsprozessmodelle, stellen Jiménez-Ramírez et. Al die Möglichkeit vor, User Feedback während der Ausführung zu erfassen indem ein automatisiert ein Fragebogen erstellt wird. Mithilfe der Beantwortung des Fragebogens wird die für die aktuelle Instanz passende Variante vorgeschlagen.

Lübke & Knaus [LK07] beschreiben einen Ansatz, bei dem User Feedback über ein sogenanntes “Experience Forum” erfasst wird. Geschäftsprozesse und deren unterstützenden IT-Systeme können auf Basis des erfassten Feedbacks optimiert werden. Zusätzlich wird das User Feedback einzelner User allen anderen Beteiligten als Hilfe während der Ausführung der Aktivitäten angezeigt. Der Ansatz von Laguna & Marklund beschreibt den „umgekehrten Weg“ von User Feedback während der Analyse- und Optimierungsphase. Auszuführende Personen (User) erhalten auf Basis von Messungen (z. B. Kennzahlenanalyse) und weiteren Erfahrungen (z. B. Kundenreviews) ein Feedback des Geschäftsprozessverantwortlichen [LM13]. Aus Basis dieses Feedbacks sollen die Ausführungen der Aktivitäten optimiert und gleichzeitig die User besser eingebunden werden.

Aus der Sicht der Autoren wird die Phase der Ausführung/Feedback durch die existierenden Ansätze nicht ausreichend unterstützt. Die Einbindung aller Beteiligten über den kompletten Geschäftsprozessmanagementzyklus ist hingegen ein kritischer Erfolgsfaktor des Geschäftsprozessmanagements [MB14].

Die in Abschnitt 3 beschriebene mobile App verbessert die Einbindung der User und die Erfassung von Feedback und somit die Möglichkeiten der Optimierung von Geschäftsprozessen innerhalb des Geschäftsprozessmanagementzyklus.

2.2 Qualitätsmodell

Um zunächst abstrakte Qualitätsmerkmale wie die User Experience für bspw. Software oder Geschäftsprozesse zu bestimmen, hat sich in der Vergangenheit der Aufbau eines Qualitätsmodells etabliert. Gemäß der Methode Factor Criteria Metric (FCM) von McCall [MRW77] wird das zu ermittelnde Qualitätsmerkmal schrittweise in weitere Qualitätsmerkmale und schließlich in Qualitätsindikatoren verfeinert. Im Vergleich zu Qualitätsmerkmalen können Qualitätsindikatoren auf Basis vorliegender Informationen konkret bestimmt werden. Sie geben dabei einen Hinweis, repräsentieren somit einen Indikator, für die Ausprägung des übergeordneten Qualitätsmerkmals. Qualitätsindikatoren können zusätzlich formal in Form einer Metrik beschrieben werden, wodurch eine Quantifizierung erfolgen kann.

In der Vergangenheit wurden für verschiedenste Bereiche Qualitätsmodelle entwickelt. Ein Beispiel, welches die verschiedenen Stufen vollständig abbildet, stellt das Qualitätsmodell für den Entwurf von Services in serviceorientierten Architekturen (SOA) von Gebhart [Ge11] dar. Dieses Qualitätsmodell zeigt, dass sich die Qualität eines Entwurfs von Services aus verschiedenen Qualitätsmerkmalen wie bspw. der losen Kopplung und Autonomie zusammensetzt. Diese Qualitätsmerkmale können weiter in Qualitätsindikatoren verfeinert werden. Beispiele sind die Asynchronität und die Abhängigkeiten von Services untereinander. Beispiele für die Formalisierung der Qualitätsindikatoren in Form von Metriken werden von Gebhart et al. in [GA11] gegeben. Diese ermöglichen basierend auf konkreten Artefakten eine Quantifizierung der Qualitätsindikatoren und somit eine effiziente Messung. Jedoch hat bereits dieses Qualitätsmodell gezeigt, dass viele Qualitätsindikatoren nur ermittelt werden können, wenn ein Kontextwissen wie bspw. Domänenwissen vorliegt, das über Informationen aus digitalen Artefakten wie Source Code hinausgeht.

Für Geschäftsprozesse wurden in der Vergangenheit mehrere Qualitätsmodelle erarbeitet, die jedoch immer nur einen Ausschnitt der Ebenen abbilden. Entweder werden wie von Suarez et al. [SFS11] sehr konkrete, messbare Qualitätsindikatoren vorgestellt, die jedoch keinen Bezug zu abstrakteren Qualitätsmerkmalen aufweisen oder es werden Qualitätsmerkmale präsentiert, die jedoch aufgrund fehlender Qualitätsindikatoren nicht ermittelt werden können.

Für letzteres ist das Qualitätsmodell von Lohrmann und Reichert [LR13] ein Beispiel. Dieses verfeinert den abstrakten Begriff der Qualität von Geschäftsprozessen in mehrere Qualitätsmerkmale unterteilt in Wirksamkeit (engl. efficacy) und Effizienz (engl. efficiency). Hierdurch wird deutlich, aus welchen unterschiedlichen Facetten die Qualität von Geschäftsprozessen besteht.

Ein Bestandteil hiervon ist die formale und informale Dokumentation von Geschäftszielen. Diese beschreibt die von außen ersichtliche Qualität des Geschäftsprozesses mit Bezug zu dessen Umwelt, jedoch wird dieses Qualitätsmerkmal nicht weiter beleuchtet. In Gebhart et al. [GMW14] wurde daher gezeigt, wie diese Qualitätsmerkmale weiter verfeinert und somit ein Qualitätsmodell für Geschäftsprozesse erstellt werden kann, welches die (teil-)automatisierte Analyse von Prozessmodellen ermöglicht. Ähnlich wie im Qualitätsmodell für serviceorientierte Architekturen, wurde jedoch auch hier deutlich, dass nicht alle Qualitätsmerkmale vollständig automatisiert ermittelt werden können, sondern Expertenwissen erforderlich ist, was eine Interaktion erfordert.

Während sich die Qualitätsmodelle für Geschäftsprozesse primär der Qualität von Prozessmodellen und harten Qualitätsmerkmalen widmen, die sich aus der Ausführung von Geschäftsprozessen durch bspw. Workflow-Engines ableiten lassen, bleiben die weichen Qualitätsmerkmalen, die die Zufriedenheit der Anwender betreffen weitestgehend außen vor. In der Softwareentwicklung hat sich dieser Aspekte bereits als User Experience etabliert. Unter User Experience ist gemäß ISO 9241-210 "A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service" zu verstehen. Der Standard gibt dabei umfangreiche Auskunft darüber, woraus die User Experience besteht. Aus diesem Grund eignet sich der Standard, um mögliche Qualitätsmerkmale und -indikatoren für Geschäftsprozesse abzuleiten, die sich mittels User Feedback während der Ausführung ermitteln lassen. So umfasst User Experience die Usability und erweitert diese um Aspekte, die den Zeitpunkt vor und nach der Nutzung des Produkts oder Systems betreffen. Aus diesem Grund können auch Qualitätsmerkmale und -indikatoren der Usability und somit der gesamte Standard der ISO 9241 in die Bewertung der User Experience mit einbezogen werden.

ISO 9241 beschreibt, dass die Arbeitsumgebung einen wesentlichen Bestandteil für das Wohlbefinden des Anwenders darstellt. Hieraus lässt sich somit ableiten, dass im Rahmen des User Feedbacks die Zufriedenheit mit den Arbeitsmitteln, dem Arbeitsplatz und der psychischen und sozialen Umgebung ermittelt werden sollte. Das Qualitätsmerkmal der psychischen und sozialen Umgebung kann wiederum in den Qualitätsindikator „Stressempfinden“ verfeinert werden, der durch den Anwender verstanden und eingeschätzt werden kann. Die folgende Abbildung zeigt den Ausschnitt des Qualitätsmodells.

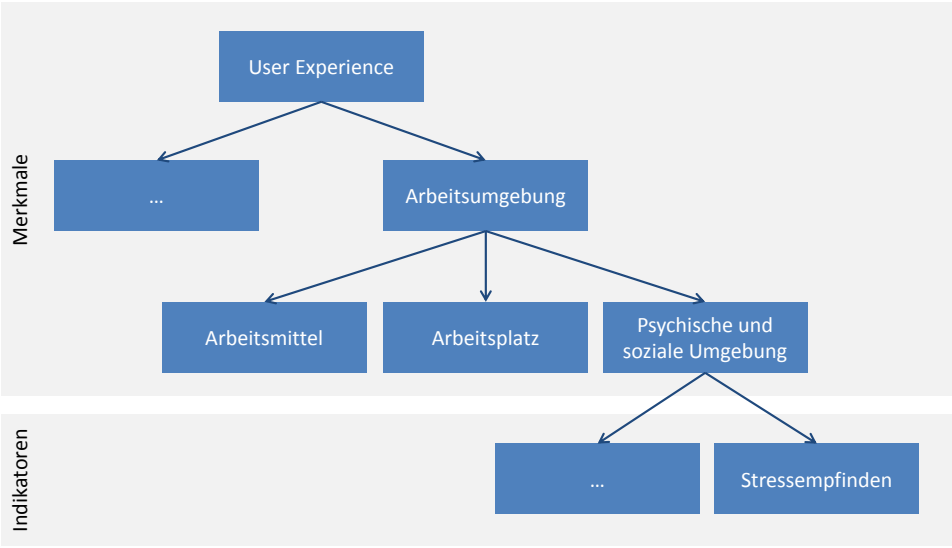


Abbildung 2: Auszug eines Qualitätsmodells zur Ermittlung der User Experience

3 Konzeption und Implementierung der mobilen App

In diesem Abschnitt wird die Architektur und Implementierung der mobilen App vorgestellt. Zu Erstellung der mobilen App wurde das BPM-System Axon.ivy eingesetzt [AX14]. Mit Axon.ivy können Geschäftsprozesse modelliert, automatisiert und überwacht werden. Des Weiteren kann Axon.ivy als Entwicklungsplattform von individuellen BPM Applikationen, wie beispielsweise der mobilen App zur Erfassung von Echtzeit User Feedback, genutzt werden.

3.1 Ansatz

Existierende (mobile) Apps fokussieren vorrangig die Definition von SOLL-Kennzahlen oder die Auswertung von „harten“ IST-Werten. Mit Hilfe der entwickelten mobilen App können neben diesen SOLL und IST-Daten (z.B. SOLL/IST-Durchlaufzeit einer Aktivität) weitere Informationen erfasst werden.

Die Abbildung 3 stellt ein Beispiel für die einzelnen Erfassungszeitpunkte an der Aktivität „Angebot erstellen“ dar:

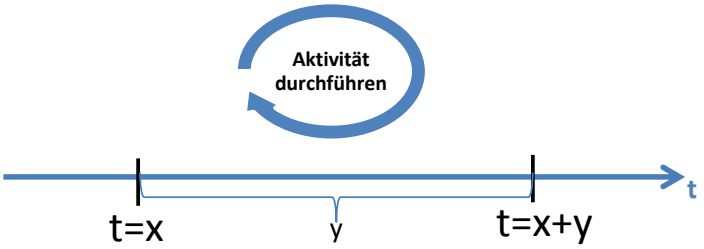


Abb. 3: Erfassung von Daten während der Ausführung einer Aktivität

Im Beispiel wird ein Geschäftsprozess zur Angebotserstellung ausgeführt (vgl. Abbildung 3): Der Zeitpunkt t=x befindet sich unmittelbar „vor“ der auszuführenden Aktivität „Angebot erstellen“. Die vordefinierten SOLL-Kennzahlen und Qualitätskriterien liegen zu diesem Zeitpunkt bereits fest und der Zeitpunkt t=x dient als Startpunkt. Parallel - während der Ausführung - werden in Zeitspanne y zu den vordefinierten Qualitätskriterien, mit Hilfe von mobil angezeigten Dialogen, User Feedback erfasst und gespeichert. Beispielsweise können User (also Anwender die den Geschäftsprozess real ausführen), Feedback zur IST-Situation der sozialen und physischen Umgebung (z.B. sehr laute Umgebungsgeräusche im Großraumbüro) mitteilen. Zum Zeitpunkt t=x+y werden im Allgemeinen weitere Messungen durchgeführt (z.B. IST-Durchlaufzeit der ausgeführten Aktivität). Die anschließende Analyse von SOLL, IST –und Feedbackdaten ermöglichen einen detaillierten „Einblick“ in die Ausführung von Geschäftsprozessaktivitäten und bilden somit die Basis für Optimierungen von Geschäftsprozessmodellen.

3.2 Architektur

Die mobile App basiert auf einer dreischichtigen Architektur. Die oberste Schicht dient der Integration des Users. Die Benutzeroberfläche ist hierbei sowohl für eine Desktop- als auch Mobildarstellung optimiert und unabhängig vom eingesetzten Endgerät aufgerufen werden. Die mittlere Ebene beinhaltet die Logik zur Ausführung der Geschäftsprozessmodelle, die Logik zur Erfassung des User Feedbacks und die Verwaltung der Qualitätskriterien. Die Ausführungslogik wird hierbei mit Hilfe einer BPM Engine implementiert. Abbildung 4 skizziert die Architektur:

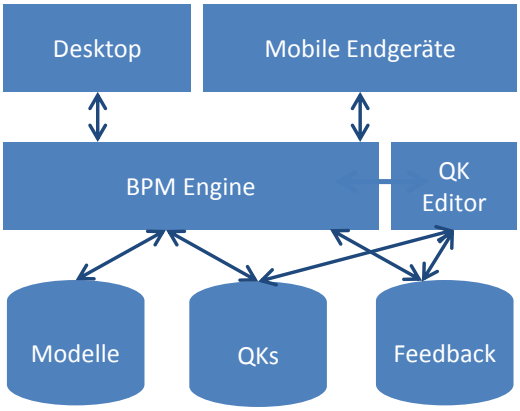


Abb. 4: Skizze der "Mobilen App" Systemarchitektur

Des Weiteren dient ein separates User Interface (OK Editor) zur Pflege der Qualitätskriterien. Durch die Trennung von BPM Engine und QK Editor können zum Beispiel unterschiedliche BPM Engines integriert werden.

Die unterste Ebene dient zur Speicherung der Daten, wobei drei getrennte Datenbanken zur Speicherung der Daten eingesetzt werden. Hierdurch wird gewährleistet, dass die Modelle, Qualitätskriterien und Feedback unabhängig verwaltet werden kann.

3.3 Implementierung

Die Implementierung der mobilen App wurde mit Hilfe der Axon.ivy BPM durchgeführt. Die Implementierung von Geschäftsprozessapplikationen wird in drei Teile gegliedert und in Axon.ivy nacheinander umgesetzt: a) Erfassung der Modelle, b) „Integration Mensch“ durch Erstellung von (mobilen) Benutzeroberflächen, c) „Integration Maschine“ durch die Anbindung von Drittsystemen. Die Abbildung 5 zeigt einen Screenshot aus der Axon.ivy BPM Suite.

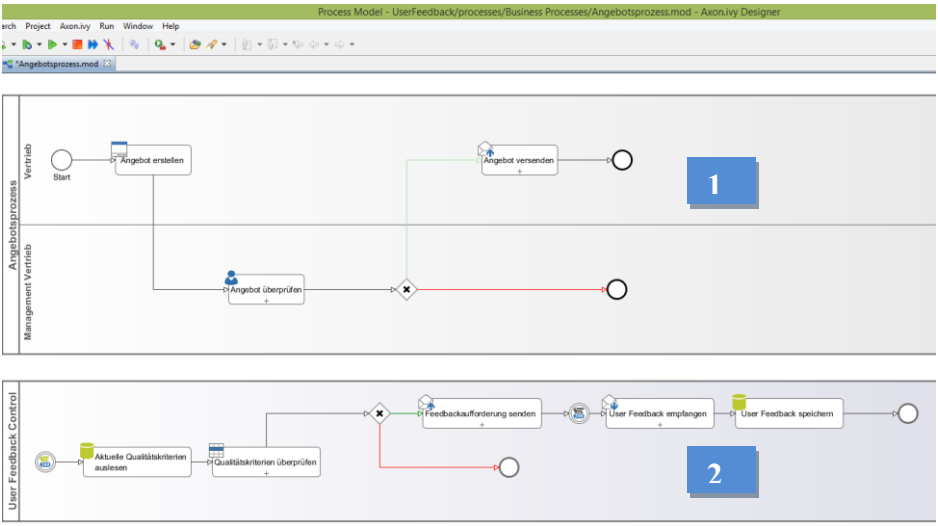


Abb. 5: Darstellung der Geschäftsprozessapplikation in Axon.ivy

Zum einen zeigt die Abbildung 5 einen Teil des Geschäftsprozessmodells „Angebotserfassung“ (vgl. (1)). Zum anderen beinhaltet die Axon.ivy App einen generischen Teil (vgl. (2)). Dieser Teil ist über den Axon.ivy Anwendungskontext mit den überliegenden Geschäftsprozessen verbunden und triggert auf Basis von vordefinierten Events die Erfassung von User Feedback (z.B. beim Start einer Aktivität). Die Einbindung des generischen Teils erfolgt über einen Komponentenansatz. Durch die Kapselung der Feedback-App in eine Komponente, kann diese flexibel in beliebige Geschäftsprozessapplikationen eingebunden werden.

4 Szenario und Evaluation

In diesem Abschnitt wird der Einsatz der mobilen App an einem Szenario beschrieben und evaluiert.

4.1 Szenario

Die mobile App wird im Rahmen der BPM(N)Easy-Vorgehensweise (vgl. [MW13] [MSW13] [MOW14]) eingesetzt. Mit BPM(N)Easy wird der Geschäftsprozessmanagementzyklus agil durchlaufen, sodass der Ansatz insbesondere für agile Umgebungen (z.B. häufige Anpassungen von Geschäftsprozessen) geeignet ist. Zusätzlich liegt der Fokus von BPM(N)Easy auf der Einbindung aller

Geschäftsprozessbeteiligten.

Durch kurze Iterationen (ein kompletter Durchlauf durch den Zyklus) werden die kontinuierliche Verbesserung der Geschäftsprozessmodelle und somit auch deren Ausführung gewährleistet. Der Einsatz einer mobilen App eignet sich hierbei besonders. Zum einen wird eine weitere Möglichkeit generiert, die flexible Einbindung von Geschäftsprozessbeteiligten (ausführender User kann direkt Feedback geben) zusätzlich zu verbessern. Zum anderen kann User Feedback während der Ausführung agil erfasst werden.

Das Szenario beschreibt einen Geschäftsprozess zur Erstellung eines Angebots. Abbildung 5 stellt einen Ausschnitt des modellierten Geschäftsprozesses dar. Innerhalb des Szenarios wird auf die Aktivität „Angebot erstellen“ fokussiert. In dieser Aktivität füllt der User ein (mobiles) Formular aus zur Erfassung von Daten (z.B. Kundenname, Beschreibung des Angebots etc.).

Folgende Qualitätsmerkmale sind für die Aktivität „Angebot erstellen“ vordefiniert:

QM Bezeichner	Kategorie	Wert	Beschreibung
Durchlaufzeit	Hart	30 Minuten	Die SOLL Bearbeitungszeit sollte 30 Minuten nicht überschreiten.
„Stressfaktor“	weich	Ja/Nein	Die Abfrage des OK „Stress“ während der Ausführung, dient zur Erfassung der User Experience (subjektives Gefühl).
„Arbeitsmittel“	weich	Ja/Nein	Sind die für die Ausführung relevanten Arbeitsmittel konkret vorhanden (z.B. Hardware, Software,…)

Tab. 1: Auflistung von Qualitätskriterien

4.2 Evaluation

Die Aktivitäten des automatisierten Geschäftsprozess wurden über ein mobiles Prozessportal verwaltet. Über das Prozessportal kann ein User authentifiziert werden (Login mit Username und Passwort). Zusätzlich erhalten die eingeloggten User eine persönliche Aufgabenliste (vgl. Abbildung 6).

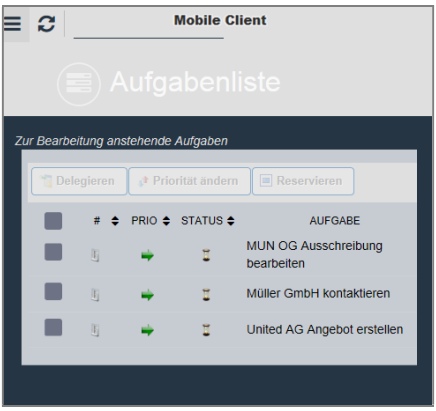


Abb. 6:Screenshot einer persönlichen Aufgabenliste in Axon.ivy

Der User kann durch einen „Touch“ auf eine Aufgabe zusätzliche Informationen anzeigen oder die Aufgabe direkt annehmen. Das Annehmen der Aufgabe triggert die damit verbundenen (technischen) Funktionen. Hat der User beispielsweise die Aufgabe eine Angebotsversendung zu genehmigen, kann der „Touch“ auf Annehmen bereits die Genehmigung und den weiteren Ablauf des automatisierten Geschäftsprozesses triggern. Im Kontext dieses Artikels wird beim Annehmen der Aufgabe dem User eine mobile Eingabemaske zur Erfassung von relevanten Daten (z.B. Kundenname, Beschreibung, …) angezeigt. Die Abbildung 7 stellt die mobilen Benutzeroberflächen zur Bearbeitung der Aufgabe „United AG Angebot erstellen“. Mithilfe der Benutzeroberfläche (vgl. Abbildung 7, linke Seite) können Geschäftsprozess-relevante Informationen eingegeben werden. Das User Interface zur Bearbeitung des Angebots ist Teil einer Axon.ivy Geschäftsprozessapplikation. Abhängig von vordefinierten Regeln erscheint (z.B. zum Start der Aufgabe), der durch die Komponenteneinbindung verfügbare, Feedback-Dialog zur Erfassung des Echtzeit-Feedbacks (vgl. Abbildung 7, rechte Seite).

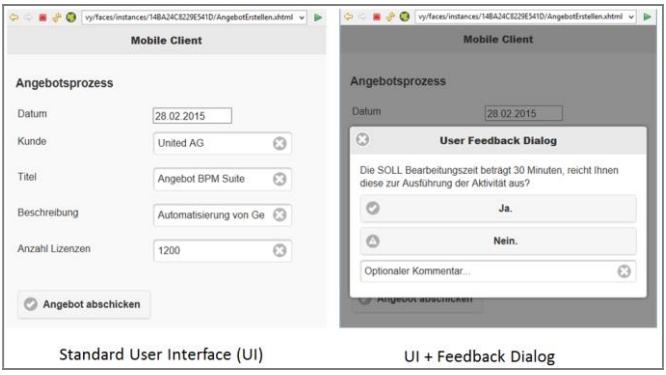


Abb 7: Screenshot der implementierten Benutzeroberflächen

5 Zusammenfassung und Ausblick

In diesem Artikel wurde ein Ansatz zur mobilen Erfassung von User Feedback vorgestellt. In Abschnitt 2 wurden existierende Arbeiten beschrieben und deren Defizite im Kontext der Echtzeit-User Feedback Erfassung identifiziert. Zur Erweiterung bestehender Ansätze, wurden in diesem Artikel sowohl weiche als auch harte Qualitätskriterien beschrieben. Basierend auf der Axon.ivy BPM Suite wurde anschließend eine mobile App implementiert, welche auf Grundlage dieser Qualitätskriterien die Erfassung von User Feedback ermöglicht. Hierbei teilt sich die mobile App in zwei Bereiche auf: (1) das Frontend bestehend aus mobilen Dialogen, welche beim Aufruf der auszuführenden Aktivität zur Erfassung des User Feedbacks dienen. (2) das generisch einsetzbare Backend, welches die Qualitätskriterien, Trigger (wann ein Dialog angezeigt werden soll) und das User Feedback verwaltet. Zur Evaluation des Ansatzes wurde ein Szenario beschrieben, in dem User während der Ausführung Auskunft zu Ihrer Bearbeitungszeit abgeben. Die Analyse der Ergebnisse verdeutlicht, dass die Optimierung von Geschäftsprozessmodellen signifikant verbessert werden kann, sofern neben den SOLL- und IST-Daten auch User Feedback erfasst wird.

Als nächster Schritt wird die technische Weiterentwicklung (z.B. Optimierung der Benutzeroberflächen) der App fokussiert. Beispielsweise soll in einer weiteren Version der mobilen App User Feedback nicht nur textuell (in grafischen User Dialogen), sondern auch via Audio- oder Videosequenz erfassbar sein. Ergänzend dazu wird analysiert, zu welchen Zeitpunkten bestmöglich User Feedback während der Ausführung von Aktivitäten erfasst werden kann. Des Weiteren wird untersucht werden inwieweit User Feedback auch zur automatischen Qualitätsoptimierung und Generierung von Geschäftsprozessmodellen genutzt werden kann.

6 Literaturverzeichnis

[AM11] Antunes, P.; Mourão, H.: Resilient Business Process Management: Framework and services. In Expert Systems with Applications, 2011, 38; S. 1241–1254.

[ATW03] Aalst, Wil van der; Ter Hofstede, A.; Weske, M.: Business process management: A Survey. International conference, BPM 2003, Eindhoven, the Netherlands, June 26-27, 2003 proceedings. Springer, Berlin, New York, 2003.

[AX14] AXON IVY AG: Axon.ivy BPM Suite. http://www.axonivy.com/de-de/, 10.04.2015.

[ba14] bamero: Bamero - Erfassung von Geschäftsprozessen The first mile in BPM. http://bamero.de/de/, 26.03.2015.

[GA11] Gebhart, M.; Abeck, S.: Metrics for Evaluating Service Designs based on SoaML. In: International Journal on Advances in Software, 4(1&2), 2011, S. 61-75.

[Ge11] Gebhart, M: Qualitätsorientierter Entwurf von Anwendungsdiensten. KIT Scientific Publishing, 2011. ISBN 978-3-86644704-2.

[GMW14] Gebhart, M.; Mevius, M.; Wiedmann, P.: Application of Business Process Quality Models in Agile Business Process Management, eKNOW 2014, 2014.

[BR10] Brocke vom, J.; Rosemann, M.: The Six Core Elements of Business Process Management. In (Vom Brocke, J.; Rosemann, M. Hrsg.): Handbook on business process management. Springer, Berlin, London, 2010, S.107-122.

[In14] Initiative D21: Digital-Index 2014.

[Ji14] Jiménez-Ramírez, A. et al.: Automatic Generation of Questionnaires for Supporting Users during the Execution of Declarative Business Process Models. In (van der Aalst, W. et al. Hrsg.): Business Information Systems. Springer International Publishing, Cham, 2014; S. 146–158.

[KO15] Koschmider, A.; Oberweis, A.: Recommendation-Based Business Processes Design. In (Vom Brocke, J.; Rosemann, M. Hrsg.): Handbook on Business Process Management 1. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015; S. 323–336.

[LK07] Lübke, D.; Knauss, E.: Dealing with User Requirements and Feedback in SOA Projects: Proceedings of Workshop on Software Engineering Methods in Service Oriented Architecture, Hannover, 2007.

[LR13] Lohrmann, M.; Reichert, M.: Understanding Business Process Quality. In: Business Process Management – Theory and Applications, Springer, 2013.

[LM13] Laguna, M.; Marklund, J.: Business process modeling, simulation, and design, 2013.

[MB14] M.E. de Waal, Benny; Batenburg, R.: The process and structure of user participation: a BPM system implementation case study. In Business Process Management Journal, 2014, 20; S. 107–128.

[Me06] Mevius, M.: Kennzahlenbasiertes Management von Geschäftsprozessen mit Petri-Netzen. Verl. Dr. Hut, München, 2006.

[MOW14] Mevius, M.; Ortner, E.; Wiedmann, P.: Gebrauchssprachliche Modellierung als Grundlage für agiles Geschäftsprozessmanagement. In (Fill, H.-G. et al. Hrsg.): GI-Edition Proceedings Band 225 - Modellierung 2014 -. 19.-21. März 2014 in Wien. Köllen, Bonn, 2014; S. 169–184.

[MOW15] Mevius, M.; Ortner, E.; Wiedmann, P.: Enhanced Stakeholder Socialization using Common Language in Agile BPM - Living business processes models instead of rigid documentations: eKNOW 2015, 2015.

[MRW77] McCall, J.; Richards, P.; Walters, G.: Factors in Software Quality, Vol. 1, 1977.

[MSW13] Mevius, M.; Stephan, R.; Wiedmann, P.: Innovative Approach for Agile BPM: eKNOW 2013, 2013; S. 160–165.

[MW13] Mevius, M.; Wiedmann, P.: BPM(N)Easy 1.2 – Gebrauchssprachliche Gestaltung IT-basierter Prozesse. In (Schmietendorf, A.; Hanin, M. Hrsg.): BSOA/BCloud 2013. 8. Workshop Bewertungsaspekte service- und cloudbasierter Architekturen, 12. November 2013, Basel/Schweiz. Shaker, Herzogenrath, 2013; S. 31–46.

[Ob96] Oberweis, A.: Modellierung betrieblicher Abläufe. In (Oberweis, A. Hrsg.): Modellierung und Ausführung von Workflows mit Petri-Netzen. Vieweg+Teubner Verlag, Wiesbaden, 1996; S. 31–51.

[OM11] OMG: Documents Associated with Business Process Model and Notation (BPMN) Version 2.0. http://www.omg.org/spec/BPMN/2.0/, 30.03.2015.

[Te13] Telefónica Germany GmbH & Co. OHG: Telefónica Deutschland _10 Trendstudie - Mein Digitales Ich - Leben in der Netzwerkgesellschaft, 2013.

[ULM15] Unger, M.; Leopold, H.; Mendling, J.: How much Flexibility is good for Knowledge Intensive Business Processes: A Study of the Effects of Informal Work Practices: 48th Hawaii International Conference on System Sciences, 2015; S. 4990–4999.

[We12] Weske, M.: Business process management. Concepts, languages, architectures. Springer, Berlin, New York, 2012.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

From Business Process Models to Capability Models

Hasan Koç[1], Kurt Sandkuhl[2]

Abstract: Enterprises are confronted with rapidly changing situations in regulations, globalisation, time-to-market pressures and advances in the technology. Management and design of capabilities allowing for more flexibility in business services is one way to tackle these challenges. This requires a shift from a business process oriented to capability management oriented view. In this respect this work introduces a business process based capability design strategy, which is a part of a lately proposed capability driven development (CDD) approach. The main contributions of this paper are i) an exploration of the role of business processes in capability modelling, ii) a component-wise structured capability development method based on business processes of an enterprise and iii) a demonstration of the method application in a use case from the utilities industry.

Keywords: BPM, CaaS, Capability Design, Context Modelling, Enterprise Modelling, Capability Driven Development, Method Engineering.

1 Introduction

In a globalized economy, many enterprises are confronted with a turbulent market environment due to rapid changes in regulations, customer demands, time-to-market pressures as well as the advances in technology. For a sustainable competitive advantage enterprises need to adjust their offerings to the dynamically changing circumstances [TPS97]. As a contribution to tackle these challenges and to offer flexible and agile business services capability management and design is considered as a promising approach. In this area the EU-FP7 project "Capability as a Service" (CaaS) develops the Capability Driven Development (CDD) approach that envisions customization of business services on the basis of capabilities and adjustment delivery according to the current context [Bē15].

Business Process Management (BPM) is of growing importance in both IS practice and research. Recently, improving business processes was named the number one priority for CIOs worldwide [NPS11]. CaaS approach aims capturing the factors that are decisive for flexibility, dynamics and variability in business services, which are reflected in business process models in organisations. In order to implement capability management, we argue for the need to shift from a business process oriented view to capability management oriented view, which requires modelling the application context of the business service

[1] University of Rostock, Institute of Computer Science, Chair of Business Information Systems, Albert-Einstein-Straße 22, hasan.koc@uni-rostock.de
[2] University of Rostock, Institute of Computer Science, Chair of Business Information Systems, Albert-Einstein-Straße 22, kurt.sandkuhl@uni-rostock.de

explicitly. The main contributions of this paper are i) exploring the role of business processes in capability modelling, ii) a component wise structured capability development method based on business processes of an enterprise and iii) a demonstration of the method application in a use case from the utilities industry.

The remaining part of the paper is structured as follows: Section 2 discusses related areas. Then section 3 introduces the CaaS Paradigm, compares it with business process as a service and proposes a business process based method for capability design. Following, section 4 illustrates the application of the method in utilities industry. Finally, section 5 summarizes the findings and concludes the work.

2 Related Work

The most relevant areas of related work for this paper are capability modelling approaches (section 2.1) and the role of business process management in capability design (section 2.2). These areas are briefly discussed in this section.

2.1 Approaches in Capability Modelling

Capability is a widely used term in the information systems (IS) literature. Although there seems to be an agreement about what constitutes a capability, it is hard to find a standard definition. The definitions mainly put the focus on "combination of resources", "capacity to execute an activity", "perform better than competitors" and "possessed ability" [SKS14].

A general consensus is that the capabilities are enablers of competitive advantage; they help companies to continuously deliver a certain business value in dynamically changing circumstances [St12]. According to [CT12] performance of an enterprise is the best, when the enterprise maps its capabilities to IT applications. Capabilities as such are directly related to the provision of business processes as a service, which are affected from the changes in the application context, such as, regulations, customer preferences and system performance. As service consumers act in rapidly changing environments, the delivery models should be extended with an additional layer that allows for anticipating variations and responding to them. In other words, adaptations to changes in delivery context can be realized promptly if the required variations to the standard processes have been anticipated and defined in advance.

An ongoing EU FP7 project "Capability as a Service in Digital Enterprises" (CaaS) received a lot of attention that envisions context-aware design and analysis of IS using the capability notion. In line with CaaS approach, capability is defined in this work as *the ability and capacity that enable an enterprise to achieve a business goal in a certain context [Bē15]*. Cornerstones of CaaS approach is described in Section 3.1. in detail.

2.2 BPM in Capability Design and Development

Business processes are parts of larger grained business services, which are usually developed for a defined customer group. However, service delivered to one specific customer from this group still has to be adapted due to rapid changes in the business environment. Since the business processes represented as process models underlie the offered business services offered, the service adaptation requires configuration and adjustment of the business processes at runtime level.

In order to ease adaptation of business services to new delivery contexts CaaS approach uses the notion of capabilities, which links business goals to business processes and actors. For this purposes, there is a need to study or develop methods that explicitly helps to define (a) the potential delivery context of a business processes, (b) the potential variants of the business processes for the delivery context and (c) what aspect of the delivery context would require what kind of variation or adaptation of the business process at hand. All three aspects require design and analysis of capabilities with particular focus on business process models.

The way methods are applied within CaaS is an extension of the method conceptualization proposed by Goldkuhl et al. [GLS98] consisting of concepts, activities and a notation. The concepts specify what aspects of reality are regarded as relevant in the modelling process, activities describe in concrete terms how to identify the relevant concepts in a method component and the notation specifies how the result of the procedure should be documented.

We conducted a systematic mapping study by analyzing 112 journals and 24 conference proceedings on the methods for design and development of capabilities. One part of the mapping study observed amongst others role of BPM in the field of capability management and exposed a stable course over the observed 15 years. However, the works put the emphasis mostly on the maturity models assessing and evaluating the quality of business processes. In the contrary, only few articles propose methods that support development of capabilities with a specific attention on business processes. [NPS11] and [Or12] investigate BPM topic from the Dynamic Capability point of view and present a framework, which supports the design of BPM capabilities. The framework consists of three activities, namely sensing, seizing and transformation, which are further elaborated in sub-capabilities. [Ad09] offers a method for IT capability based business process design, which consists of 8 steps. The works described provide a good starting point to design capabilities based on business processes but still cannot be taken solely as capability development methods in line with the method conceptualization of Goldkuhl [GLS98]. The two main reasons behind this is i) the lack of important concepts and outputs of the activities in the proposed approaches and ii) loosely description of the roles and notations, which are used to represent the concepts. As a result, we identified the need to engineer a capability development method based on the business services of an organisation, which are in designed and modelled as business processes.

3 From Business Process Models to Capability Models

3.1 Capability as a Service – A Novel Paradigm

Operating in a changing environment increases the importance of business agility in terms of customization, adaptation and adjustments. Organisations need to take contextual variations such as local legislations, location, work schedules and changes in regulations into account when offering their services. In this respective CaaS project aims to design the business services responding to the change needs in accordance with the current context. To this end, a Capability Driven Development (CDD) approach is being developed which enables organisations to design services based on the notion of capability that are adjusted according to the current context. CDD consists of both a methodology, comprising enterprise modelling, capability design and capability delivery, a suite of tools, e.g. a capability modelling tool, a platform for context monitoring, and a Capability Navigation Application that calculates indicators and performs automatic runtime adjustments [SKS14].

The CDD is based on the capability meta-model (CMM) presented in Fig. 1, which is developed on the basis of industrial requirements and related research on capabilities [St12], [Zd13] and evaluated in numerous use cases [Es14]. A *capability* is defined by specific business services represented as business process models, an application context for these business services and goals of the enterprise to be reached. They are required by enterprise *goals* and their delivery is supported by *patterns*, which realize runtime adjustments of the capability in the business context. The business context is represented by a number of concepts. *Context set*, which is basically set of *context element*s including their ranges, defines the relevant context for capability delivery. The permitted values of the context elements are specified in *context element range* class and measured by *measurable property* class. Context elements are related to different business drivers that cause variations in processes. *Variation aspect* identifies the causes and types of such variations and *variation point* locates the positions of the variations in the business service model. Process variants are adjustments of a master process (e.g., a reference process) to specific requirements building the process context [Rv07].

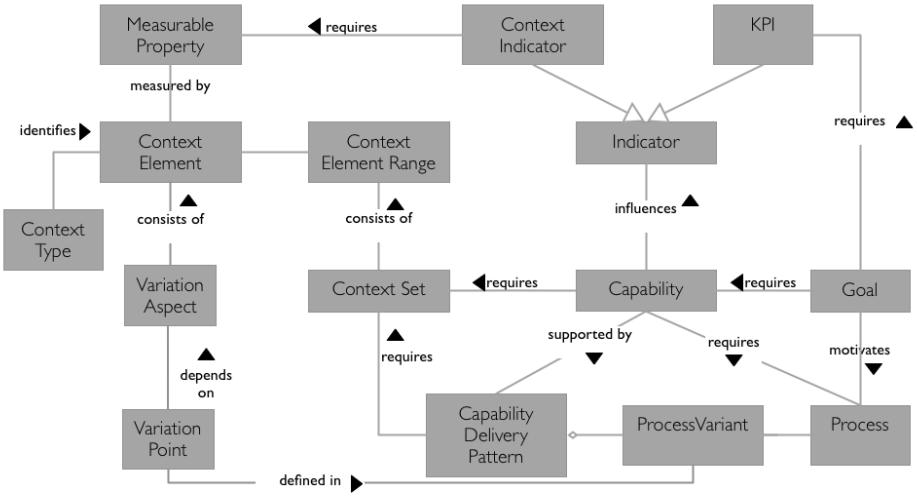


Fig. 1: Simplified view of CDD Meta model based on [Zd13]

3.2 A Business Process Based Method for Capability Design

The CDD applies modular approach to methodology engineering and documentation by dividing the methodology into several method components. In doing so the method user could focus on those parts of the method that are needed and select the components relevant for a specific tasks "on demand" from a repository. Hence, the method is composed of three method components, namely "define scope, develop/ update enterprise models" and "context modelling".

The business process based capability modelling method proposes that the starting point of the capability design is a process underlying a business service. The business service is further refined and extended by adding context awareness and adaptability, so as to establish a capability that can deliver this service in varying circumstances. Many organisations at this level have already defined and modelled business processes that are implemented to offer business services. Hence, the method assumes that the digital enterprise that aims to offer capabilities has services modelled and implemented as business process models. The method consists of the following components.

Component 1: *Define scope.* The organisation offers services based on business processes that are already modelled. In order to design the capabilities by means of business processes the capability designer first selects the service and sets the scope of the capability design. The selection can depend on various factors, such as optimising the services with high process costs or managing services that frequently change and require the adjustment of business processes. After selecting the service the define

granularity activity sets the abstraction level, at which the processes supporting the business service to be improved are identified and analysed. Different ways to describe granularity levels in business processes exist in the literature. Our method adopts the decomposition approach proposed by [Fr13], which differentiates between a main process, which does not belong to a larger process and is decomposed into sub-processes. Regarding the business goals and offered capabilities needed to reach them, the method user most probably models at main process level. Nevertheless, the main processes might be refined by sub-processes in the next activity "identify processes. For this purpose 5-policies approach proposed by [HBR10] can be applied, which describes a general strategy for identifying processes. The approach can be adjusted for the chosen granularity level in the activity before. It should be noted that capturing possible variations of the processes are not included in these activities, but are subject to the third method component, context modelling. The "Define Scope" method component uses the "Business Process Model and Notation" (BPMN) 2.0.

Component 2: *Develop/ update enterprise models.* This method component analyses the enterprise models to make sure that selected business process models are up-to-date and applies changes if required. Moreover, the capability designed should be aligned with the goals that an enterprise aims to achieve. To check if business goals are satisfied during the capability delivery, KPIs are used to measure the achievement of goals. This step analyses and updates the goal models as well as KPIs, if any exist. If no goals model is available, then they can be developed based on the guidelines proposed in [Sa14]. Since an alignment of goals is required on the business service level, method user should rather model the capability related goals and not the enterprise objectives on a general basis. The last activity relates goals, business process models, KPIs and capabilities, which is used as input in the next method component. This method component uses BPMN 2.0 and 4EM Notation [Sa14] to represent the important concepts such as goals, processes and KPIs.

Component 3: *Context Modelling.* A capability is defined by specific business services, a defined application context for these business services and goals of the enterprise to be reached. In this method the capability designer models the potential application context where the business service is supposed to be deployed. For this purpose the designer executes three activities subsequently, namely find variations, capture context element and design context by applying the BPMN 2.0 and context modelling notation as proposed in [SK14]:

Find Variations. Identifies the variability in the business process models and focuses on their possible variations. By further specifying the variability in the following activities component, the method user aims to develop a context element. For this activity, the method user requires business process models, goals model, KPIs and a defined capability as an input. The output produced is a business process model including the process variants and variation points. We propose the following guidelines to support the method user on what constitutes a process variant and how to distinguish variability from standard decisions.

- Different than a decision point, a process variant is always relevant for capability delivery. If a decision point is identified that is important for capability delivery, then the method user checks depending what the decision is met, i.e. event-based, context-based or data-based. In order to do so, the condition expression at the decision point is evaluated. Context-based resolution of the decision point indicates that the subsequent task should be modelled as a process variant.
- The decision point is represented with a standard gateway if its resolution is data- or event-based.

Capture Context Element. This activity investigates the concepts and aspects of the context by eliciting the factors that cause variations in the processes, which are identified in the *Find Variations* activity. CDD Approach defines context as any information characterizing the situation of an entity [De01]. In line with this definition we assume that characterising information as such can stem from the factors of change, since they mainly cause variations in the business process models. Additionally, the configuration of a particular process variant depends on the context, in which the process is being implemented [HBR10]. Thus a substantial analysis of change factors reflected in business process variants is required to capture a context element. The following guidelines are proposed for this kind of task:

- The influence of the identified factor must be vital for the execution of the capability to reach a goal.
- The change factor must be measurable, i.e. its value must be retrieved from an information system.
- Context element is an external influence on the process, which should not exist as a process instance or data in BPM.
- Context element causes a variation in the process, when its value changes.

In some cases it is challenging to distinguish between process variables and context elements, for which we propose the two guidelines:

- A process variable is produced during the activities of a given process and managed by the application system. On the other hand a context element or (contextual data) is an external influence on the process itself, which should not exist as a process instance or data in the system.
- A context element can act as a filter and determine which variables have to be gathered from process instances.

Design Context. This activity defines a way to categorize a context element, such as if it is of static or dynamic nature. Defining the context type might have an influence at run-time. For instance, the value of a context element, which is dynamic, can be assessed in shorter time frames than the values of a static context element. Moreover it links the capability under study to the contextual influences by creating a “container” (a context

set), including the permitted ranges of the context elements for capability delivery (context element ranges) and what attributes to measure to enable reasoning about the context elements (measurable properties). Finally, the properties of the context elements relevant to capability design (context indicators) are identified.

Fig. 2 illustrates the component wise structured capability design method, based on the business services and hence on the process models of the organisation under study. Experiences gathered during the method engineering phase are detailed in an earlier work [SK14].

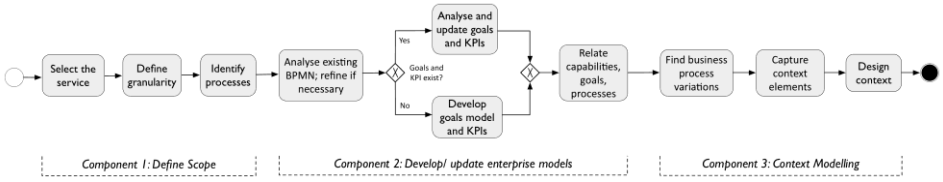


Fig. 2: A Business Process Based Method for Capability Design

4 Method Application in a Use Case

In order to illustrate the applicability of the process based capability design method introduced in section 3.2, an example use case of the industrial partner SIV.AG is described.

4.1 Use Case Description

SIV.AG from Rostock (Germany) is an independent software vendor for the utilities industry and offers different kinds of services to their clients. The target group for these services is medium-sized utility providers and other market roles of the energy sector in Germany, Bulgaria, Macedonia and several other European countries.

The company owns a business service provider (BSP) offering business process outsourcing (BPO) services, i.e. performing a complete business process for clients outside of the organization. The BSP as such provides services for the customers running kVASy®, SIV´s industry specific ERP platform. Integrated with the business process environment, the “native” kVASy® services providing business logic for the energy sector are implemented using a database-centric approach. Different deployment models are used including a provider-centric model (kVASy® and the business processes are run at SIV), a client-centric model (kVASy® is installed at the client site and the manual work of the business process is performed at SIV) and mixed models (e.g. kVASy® in the cloud, work and process performed partly at the client and partly at SIV).

In particular the BSP deals with intercompany business processes between partners in

the utility market that requires exchange of messages about energy consumption or customer master data. The exchanged messages have to be both syntactically and semantically correct before being processed further. In case of a faulty message the BSP might act as a clearing centre involving the manual interaction of a human agent. The decision whether to route a case to BSP currently depends on contextual factors, such as the policies between the BSP and the client, the available resources on both sides and the number of transactions to be cleared. This service area of BSP has to implement potential variations of the clients' way to perform business. One variation is the need to adjust the standard software systems for the organizations in question, which implements the core processes. The second cause of variation is the configuration for the country of use, i.e. the implementation of the actual regulations and bylaws. The third variation is related to the resource use for implementing the actual business process for the customer, i.e. the provision of technical and organizational capacities. Last but not least, the fourth variation is the application of the solution that remedies the faulty message, which is carried out by the knowledge worker at BSP. In this case the outsourcing business services need to be dynamically routed; i.e. it should be resolved at run-time whether the individual case should be routed to the BSP. This decision is based on run-time data such as the backlog size of the customer, the type of service supported and the type of the exception, which essentially need to be captured as a capability model. The possible variations and the simplified view of the use case is illustrated in Fig. 3.

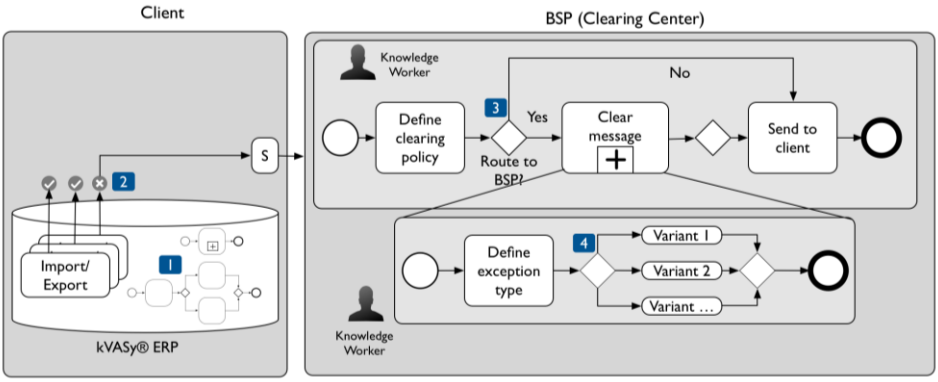


Fig. 3: Four Variation Types and a Simplified Clearing Process

4.2 Method Application in Utilities Industry

This section describes the activities needed to design capabilities of SIV.AG based on the existing business process models.

Component 1: *Define scope.* The scope of the capability delivery is to increase the throughput of energy consumption data messages (MSCONS). For this purposes "exception clearing in market communication" service is selected, as, which envisions

offering dynamic support in exchange of messages where faulty processes must be cleared by the BSP considering the agreements between SIV and the client. The granularity level is chosen as "main processes" and business processes are extracted from the process models focusing on the conditions i) whether to route the faulty MSCONS message to BSP and ii) how this message is cleared by a knowledge worker in BSP. This is shown simply on the right hand side of Fig. 3.

Component 2: *Develop/ update enterprise models.* A goals model has to be developed from the scratch in the SIV case. This required the involvement of the domain experts and product owners to the modelling sessions as well as the analysis of secondary data. Moreover, indicators are defined to measure to what extent the goals are achieved. An excerpt of the SIV´s goals model is depicted in Fig. 4, which for brevity reasons does not include KPIs. In addition to that, the business process models that are influenced by the goals are analysed, which did not require further update.

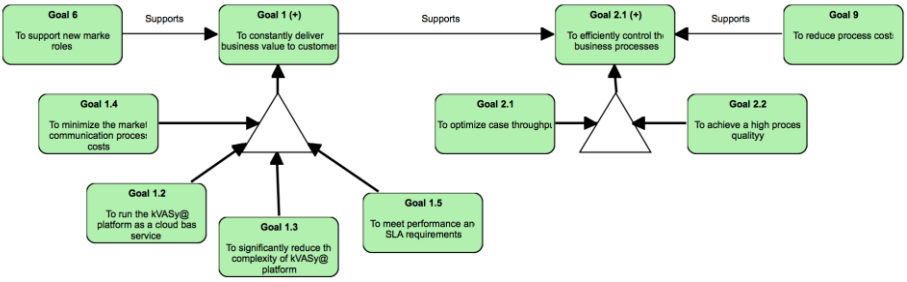


Fig. 4: SIV Goals

Component 3: *Context Modelling.* In the first activity, the variations in the business process models are analysed. We found out that depending on various aspects such as contract type and customer's market role the clearing process of a faulty message can be routed to the external BSP or to the client itself. In this respect, the business logic where to route the faulty case is represented as a variation point, followed by two variants, (Clear message/ Send to client) which is illustrated as Nr. 3. in Fig. 3. While the outcome of this variation point is just a "Yes/No" decision, the inference logic behind may be quite complex. In order to resolve this question, the knowledge worker neither analyses the data produced by the process executed at client nor meets the decision based on a single, independent event. The case is resolved by the analysis of the contextual-data, i.e. the comparison of the situation in which the faulty case was generated with the contractual agreement of the client with BSP. This contextual data is external to the process, for the capacity of the BSP is not known a priori and they are not produced during the execution of the process. Thus, next activity "Capture context element" elaborates the concept of context by studying the change factors and capturing them as "context element candidates" first.

We determined three main factors influencing the routing decision. *F1*, *clearing policy*

that the client agrees upon with BSP defines clearly by whom the exception should be cleared. This is strongly related to the dynamic support in message clearing. *F2, operating platform* is included in the contractual agreement, whereas currently its change does not necessarily have an impact on the routing decision. The final factor F3 consists of parameters, which are actually produced during process execution, such as backlog threshold of the client and the properties of the message throwing an exception. Although one can favour the treatment of such parameters as process variables, their interplay with *F1 clearing policy* is decisive for the message clearing capability. All contracts in SIV case include specifications about the message types, message versions, application references and backlog threshold. These comply with the guidelines proposed in section 3.2 and are thus identified as context elements.

The next activity Design Context specifies selected context elements. For this we defined the attributes that have to be measured in order to provide values of context elements (measurable properties) and the boundaries of permitted values for the context elements (context element ranges). After the expert evaluation we gathered the context element ranges defined for a customer in a context set, which should dynamically support the routing decisions. We benefited from defining the context elements only once, which are more or less similar in each of the clients and changing the ranges of the elements based on the clearing policy of the client. A tabular view of context set of the use case is shown in Tab. 1. It should be emphasized that currently a tool for capability modelling is being developed, which allows for the modelling of the context in line with the notation proposed in [SK14] and meta model introduced in section 3.1.

Context Element	**Context Element Range**	**Measurable Property**	**Routing Decision**
Message Type	MSCONS	UNH.S009	IF {Message Type = MSCONS} ∧ {Application Reference = VL} ∧ {Threshold = Exceeded} THEN apply Variant 2
Application Reference	VL, TL, EM	UNB.S005	
Backlog Threshold	Exceeded	Backlog Size	

Tab. 1: Context Set for SIV Case

5 Conclusion

Managing rapidly changing situations in business service offerings is a serious challenge. Today´s enterprises adjust their services on the implementation level, which requires manual intervention, such as updating the software implementation or customizing the business processes. This has three main disadvantages. First, adjusting business services for each customer´s specific application context requires an organisation's resources. Second, large numbers of business process models are

developed that usually include recurring parts and thus are variants of standard processes. Third, the meaning of models which are supposed to support stakeholder communications deteriorates since the solutions are rather designed on implementation level. In this work, we motivate the notion of capability and its design based on the business processes as a solution. A capability is defined by specific business services represented as business process models, an application context for these business services and goals of the enterprise to be reached. For this purposes we introduce a component-wise structured method that i) is motivated by the business process of organisations and thus requires a profound analysis of business process models and ii) emphasizes the modelling of the application context, where the business service is supposed to be deployed.

For the use case defined in section 4 the business process models required for capability delivery were made available to the researchers. Moreover, we executed semi-structured interviews with enterprise architects, domain experts and business service managers. Regarding our experiences on method application, we became aware of the need to differentiate between process variables and context elements, since both types of parameters originate during the execution of the process but have different design and runtime implications. For this purposes, we provided guidelines on what constitutes a context element and which parameters should be treated as process variables. These parameters were analysed in detail to decide whether they qualify as context elements or process variables. Concerning the *modelling of process variants*, the use cases have also proven that the values of such parameters have an influence on the decision logic and it is sometimes confusing to distinguish variability from business processes decisions. To solve this problem the term variation point and an initial set of guidelines were introduced.

An earlier work motivated two more capability design methods, namely the goals based and concepts based method. The goals based capability design can start with analyzing the existing goal hierarchy and/or setting new goals and then defining how they should be reached in terms of capabilities, business processes and which context properties should be considered. In concept-based method, capabilities may be designed by analyzing the existing knowledge structures and their relationships with the application context. A detailed comparison of these three different strategies can be found in [Es15].

6 Bibliography

[Ad09] Adam, S. et al.: IT Capability-Based Business Process Design through Service-Oriented Requirements Engineering. In (Halpin, T. et al. Eds.): Enterprise, Business-Process and Information Systems Modeling. Springer Berlin Heidelberg, 2009; pp. 113-125.

[An09] Anstett, T. et al.: Towards BPEL in the Cloud: Exploiting Different Delivery Models for the Execution of Business Processes: 2009 IEEE

Congress on Services (SERVICES), 2009; pp. 670–677.

[Bē15] Bērziša, S. et al.: Capability Driven Development. An Approach to Designing Digital Enterprises. In Business & Information Systems Engineering, 2015, 57; pp. 15–25.

[CT12] Chen, J.-S.; Tsou, H.-T.: Performance effects of 5IT6 capability, service process innovation, and the mediating role of customer service. In Journal of Engineering and Technology Management, 2012, 29; pp. 71–94.

[De01] Dey, A. K.: Understanding and using context. In Personal and ubiquitous computing, 2001, 5; pp. 4–7.

[Es14] España, S. et al.: Capability-Driven Development of a SOA Platform: A Case Study. In (van der Aalst, W. et al. Eds.): Advanced Information Systems Engineering Workshops. Springer International Publishing, Cham, 2014; pp. 100–111.

[Es15] España, S. et al.: Strategies for Capability Modelling: Analysis based on Initial Experiences. ASDENCA workshop at CAISE 2015. Accepted for publication in Springer LNBIP. Stockholm, Sweden, June 2015.

[Fr13] Fredrik Milani et al.: Modelling Families of Business Process Variants: A Decomposition Driven Method. In CoRR, 2013, abs/1311.1322.

[GLS98] Goldkuhl, G.; Lind, M.; Seigerroth, U.: Method integration: the need for a learning perspective. In Software, IEE Proceedings -, 1998, 145; pp. 113–118.

[HBR10] Hallerbach, A.; Bauer, T.; Reichert, M.: Configuration and Management of Process Variants. In (Vom Brocke, J.; Rosemann, M. Eds.): Handbook on Business Process Management 1. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010; pp. 237–255.

[NPS11] Niehaves, B.; Plattfaut, R.; Sarker, S.: Understanding Dynamic IS Capabilities for Effective Process Change: A Theoretical Framework and an Empirical Application. In ICIS 2011 Proceedings, 2011.

[Or12] Ortbach, K. et al.: A Dynamic Capability-Based Framework for Business Process Management: Theorizing and Empirical Application: 2012 45th Hawaii International Conference on System Sciences (HICSS), 2012; pp. 4287–4296.

[RCL09] Rimal, B. P.; Choi, E.; Lumb, I.: A Taxonomy and Survey of Cloud Computing Systems: 2009 Fifth International Joint Conference on INC, IMS and IDC, 2009; pp. 44–51.

[Rv07] Rosemann, M.; van der Aalst, W.M.P.: A configurable reference modelling language. In Information Systems, 2007, 32; pp. 1–23.

[Sa14] Sandkuhl, K.: Enterprise modeling. Tackling business challenges with the 4EM method. Springer, Heidelberg, 2014.

[SK14] Sandkuhl, K.; Koç, H.: Component-Based Method Development: An Experience Report. In (Frank, U. et al. Eds.): The Practice of Enterprise Modeling. Springer Berlin Heidelberg, 2014; pp. 164-178.

[SKS14] Sandkuhl, K.; Koç, H.; Stirna, J.: Context-Aware Business Services: Technological Support for Business and IT-Alignment. In (Abramowicz, W.; Kokkinaki, A. Eds.): Business Information Systems Workshops. Springer International Publishing, 2014; pp. 190-201.

[St12] Stirna, J. et al.: Capability Driven Development – An Approach to Support Evolving Organizations. In (Aalst, W. et al. Eds.): The Practice of Enterprise Modeling. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012; pp. 117–131.

[TPS97] Teece, D. J.; Pisano, G.; Shuen, A.: Dynamic Capabilities and Strategic Management. In Strategic Management Journal, 1997, 18; pp. p 509-533.

[Zd13] Zdravkovic, J. et al.: Modeling Business Capabilities and Context Dependent Delivery by Cloud Services. In (Hutchison, D. et al. Eds.): Advanced Information Systems Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013; pp. 369–383.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Überblick und Weiterentwicklung von Methoden zur Beherrschbarkeit von Variabilität in Geschäftsprozessmodellen

Oliver Arnold[1], Ralf Laue[2]

Abstract: Es existieren zahlreiche verschiedene Methoden und Ansätze zum richtigen Umgang mit Variabilität von Geschäftsprozessmodellen. Der vorliegende Beitrag stellt einige dieser Methoden vor und geht dabei auch auf vorhandene Vor- und Nachteile ein. Auf dieser Grundlage wird eine Reihe von Thesen formuliert, die die mögliche Weiterentwicklung dieser Methoden insbesondere im Hinblick auf ihre praktische Erprobung sowie die Einbettung in eine ganzheitliche Vorgehensweise skizzieren.

Keywords: Variabilität, Geschäftsprozessmodelle, prozedural, deklarativ, C-EPC, Provop, vBPMN, ConDec

1 Einleitung

Die Modellierung und die Analyse von Geschäftsprozessen sind wichtige Inhalte des Geschäftsprozessmanagements und unverzichtbare Grundlage für die systematische Unterstützung und Weiterentwicklung betrieblicher Abläufe. Dabei wird häufig eine Standardisierung von Geschäftsprozessen angestrebt. Gleichzeitig existieren neben einem Standardablauf oft auch zahlreiche Abweichungen und Ausnahmen aufgrund wechselnder Rahmenbedingungen, unter denen der untersuchte Prozess in der Praxis abläuft. In [Vo12] wird beispielsweise beschrieben, wie in zehn niederländischen Gemeinden die gleichen grundlegenden durch Gesetze vorgegebenen Verwaltungsprozesse auf immer wieder unterschiedliche Art und Weise in Form verschiedener Prozessvarianten umgesetzt werden.

Diese Prozess(-modell-)varianten zeichnen sich dadurch aus, dass bestimmte Teile des Prozessmodells über alle Varianten hinweg gleich bleiben, während andere Teile in den einzelnen Prozessinstanzen je nach vorliegendem Kontext variabel ablaufen. In ihrer Gesamtheit bilden solche Prozessmodellvarianten eine Geschäftsprozessmodellfamilie (GPM-Familie). Der richtige Umgang mit dieser Geschäftsprozessmodellvariabilität (GPMV) war in den letzten Jahren Gegenstand zahlreicher wissenschaftlicher Untersuchungen [Va13].

[1] Westsächsische Hochschule Zwickau, Fachgruppe Informatik, Dr.-Friedrichs-Ring 2A, 08056 Zwickau, oliver.arnold@fh-zwickau.de
[2] Westsächsische Hochschule Zwickau, Fachgruppe Informatik, Dr.-Friedrichs-Ring 2A, 08056 Zwickau, ralf.laue@fh-zwickau.de

Auch die Autoren des vorliegenden Beitrags haben 2014 einen Artikel über vorhandene Methoden zur Beherrschbarkeit der GPMV veröffentlicht [AL14a]. Wesentliche Inhalte dieses Übersichtsartikels werden im folgenden Kapitel 2 vorgestellt. In Kapitel 3 wird der Übersichtsartikel näher charakterisiert, und es werden die Ergebnisse zusammengefasst sowie auf deren Grundlage einige Thesen abgeleitet. Der Beitrag schließt mit Kapitel 4 ab, in dem der aus den Thesen resultierende Forschungsbedarf betrachtet wird.

2 Methoden zur Beherrschbarkeit von GPMV

2.1 Systematisierung der Methoden

Geschäftsprozesse werden häufig **prozedural** beschrieben, d.h. es wird die zeitlich-logische Abfolge von Aktivitäten modelliert [Pa11]. Eine Alternative dazu ist die **deklarative Beschreibung**, die beispielsweise in [PA06] vorgestellt wird. In prozeduralen Ansätzen wird anhand eines vorgegebenen Ablaufs genau beschrieben, *wie* das Ziel des Geschäftsprozesses zu erreichen ist. Deklarative Ansätze hingegen beschreiben nur, *was* getan werden soll. Beispielsweise würde in einer deklarativen Beschreibung nur festgelegt, dass zwei Aktivitäten nicht gemeinsam in der gleichen Prozessinstanz ausgeführt werden dürfen, ohne wie in einer prozeduralen Beschreibung an einer (ggf. künstlich geschaffenen) Stelle im Prozessablauf eine Entscheidung zwischen den beiden Aktivitäten zu modellieren. Beide Ansätze können allgemein zur Prozessmodellierung eingesetzt werden, wobei noch kein expliziter Bezug zur GPMV gegeben ist. Beispiele für konventionelle prozedurale Methoden finden sich im Abschnitt 2.2, ein Beispiel für deklarative Prozessmodellierung in Abschnitt 2.4.

In [To13] wird der Bereich der prozeduralen Methoden, die sich direkt mit GPMV beschäftigen, in zwei Untergruppen eingeteilt. Dies ist zum einen der sogenannte **Verhaltensansatz**. Hier liegt der Variantenbildung ein konfigurierbares Gesamtmodell (z.B. mit Platzhaltern) zugrunde. Alle Mitglieder einer GPM-Familie werden mit allen Gemeinsamkeiten und Unterschieden in einem einzigen Gesamtmodell abgebildet. Die Bezeichnung „Verhaltensansatz“ resultiert daraus, dass das gesamte in der GPM-Familie mögliche Verhalten in diesem Modell enthalten ist.

Die zweite Untergruppe wird als **Strukturansatz** bezeichnet. Dabei bildet ein als Basisprozess bezeichnetes Grundmodell den Ausgangspunkt. Zur Umsetzung der Variabilität werden dann Änderungsoperationen auf diesen Basisprozess angewandt, wodurch die einzelnen Prozessvarianten entstehen. Die Bezeichnung „Strukturansatz“ resultiert daraus, dass alle Modellvarianten durch schrittweise Änderung der Struktur des Modells entstehen.

Beispiele für Methoden, die dem Verhaltens- bzw. dem Strukturansatz zuzuordnen sind, werden in Abschnitt 2.3 vorgestellt.

2.2 Konventionelle Modellierungsmethoden

Wird keine speziell auf GPMV ausgerichtete Modellierungsmethode eingesetzt, werden in der Regel zwei prozedurale Ansätze zur Abbildung von Prozessvarianten verwendet.

Eine Möglichkeit stellt die Modellierung der verschiedenen Prozessvarianten in einem gemeinsamen GPM dar. Dies wird in der Literatur zur GPMV als Single-Model-Ansatz bezeichnet [KY12][HBR10]. Dabei entsteht meist ein großes und komplexes Modell mit zahlreichen Verzweigungen, die eigentlich der Unterscheidung zwischen den verschiedenen Prozessvarianten dienen. Beispielsweise könnte in einer Behörde ein bestimmter Prozess mit einem Online-Formular unterstützt werden, in einer anderen Behörde hingegen muss der entsprechende Antrag persönlich in der Behörde abgegeben werden. An einem bestimmten Punkt im Prozessablauf wird dann eine Entscheidung zwischen diesen beiden Vorgehensweisen modelliert. Die Entscheidung wird in derselben Einrichtung (zumindest solange sich die Rahmenbedingungen nicht ändern) immer wieder gleich ausfallen. Dennoch müsste im Modell die Entscheidung bei jedem Prozessdurchlauf immer wieder getroffen werden. Solche eigentlich der Unterscheidung zwischen verschiedenen Prozessvarianten dienenden Verzweigungen werden als „Variabilitätspunkte“ bezeichnet. Im Single-Model-Ansatz werden sie jedoch wie „normale“ (zur Laufzeit des Prozesses zu treffende)) Entscheidungen behandelt [HBR09].

Die zweite Möglichkeit wäre die Modellierung der verschiedenen Prozessvarianten in je einem eigenen separaten Modell. Dies wird in der einschlägigen Literatur als Multi-Model-Ansatz bezeichnet [KY12][HBR10]. Dabei wird das Prozessmodell kopiert und manuell angepasst. Da dennoch viele Teile des Prozessmodells in allen oder zumindest in vielen Varianten vorhanden sind, entsteht bei dieser Art der Prozessmodellierung viel Redundanz. Eine Kopplung der einzelnen Modelle ist häufig nur über die Namensgleichheit der Bezeichner gegeben. Oft erfolgt mit der Zeit eine Degeneration der Modelle, da Verbesserungen dann ggf. nur an einzelnen Varianten vorgenommen werden [HBR09].

Beide konventionellen Ansätze führen somit zu unbefriedigenden Ergebnissen.

2.3 GPMV-spezifische prozedurale Methoden

In [RA07] wird der Modellierungsansatz der Configurable Event-Driven Process Chain (C-EPC) beschrieben. Er erweitert die weitverbreitete Modellierungsmethode der Ereignisgesteuerten Prozesskette um konfigurierbare Elemente. Diese konfigurierbaren Elemente dienen der Umsetzung der Variationspunkte. Das so entstehende konfigurierbare Referenzmodell beinhaltet demnach sowohl die Gemeinsamkeiten als auch die Unterschiede der GPM-Familie. Alle zugehörigen Informationen sind somit in einem einzigen Artefakt (dem konfigurierbaren Referenzmodell) enthalten. Damit ist die C-EPC ein Vertreter des bereits vorgestellten Verhaltensansatzes.

In C-EPCs sind sowohl Funktionen als auch logische Konnektoren konfigurierbar. Konfigurierbare Funktionen können entweder für die entstehende Variante eingeschlossen (im Modell dargestellt durch „ON“), immer übersprungen („OFF“) oder nur unter bestimmten Bedingungen übersprungen („OPT“) werden.

Konfigurierbare Konnektoren können auf einen konkreten Konnektortyp abgebildet werden. Dabei kann durch Konfiguration aber nur eine weitere Einschränkung des möglichen Verhaltens vorgenommen werden. Beispielsweise könnte ein konfigurierbarer XOR- bzw. ODER-Konnektor derart konfiguriert werden, dass in der resultierenden Variante immer genau eine der nachfolgenden Sequenzen ausgeführt wird (d.h. die Auswahl wird bereits zur Konfigurationszeit getroffen). Andererseits kann ein AND-Konnektor nicht weiter eingeschränkt werden. Er verlangt immer, dass alle ausgehenden Kontrollflüsse ausgeführt werden.

Durch logische Ausdrücke können auch Abhängigkeiten zwischen den zur Verfügung stehenden Konfigurationsoptionen ausgedrückt werden.

Der Vorteil der C-EPC-Methode im Vergleich zur auf den ersten Blick ähnlich anmutenden Single-Model-Methode ist, dass Konfigurationsentscheidungen deutlich von „normalen“ Entscheidungen im Prozessablauf unterschieden werden und dass diese Konfigurationsentscheidungen bereits vorab getroffen werden, indem durch Konfiguration des Referenzmodells eine spezifische Variante erstellt wird.

In [HBR10] wird ein weiterer Ansatz zur Umsetzung von GPMV beschrieben. Der Ansatz trägt die Bezeichnung Provop, was für PROcess Variants by OPtions steht. Der Ansatz geht davon aus, dass jedes Prozessmodell aus einem anderen abgeleitet werden kann, indem eine Reihe von Änderungsoperationen angewendet werden. Damit ist Provop ein Vertreter des sogenannten Strukturansatzes, bei dem eine Prozessfamilie durch die schrittweise Modifizierung der Struktur einer bestimmten Prozessvariante entsteht.

Bei den grundlegenden Änderungsoperationen handelt es sich um das Einfügen (INSERT), Entfernen (DELETE) und Verschieben (MOVE) von Teilen eines GPMs sowie dem Verändern (MODIFY) von Prozesselementattributen. Die Operationen werden auf sogenannte Anpassungspunkte (engl. *adjustment points*) angewendet, die dem Eingang bzw. Ausgang eines Aktivitätsknotens bzw. eines logischen Konnektors entsprechen.

Die konkreten Änderungsoperationen (bzgl. eines konkreten Prozesses) werden zu Optionen gruppiert, die immer zusammen angewendet werden. Dadurch wird bis zu einem gewissen Grad vermieden, dass durch sehr viele sehr feingranulare Änderungen die Übersicht über die Zusammenhänge der GPM-Familie verloren geht. Außerdem können Beziehungen zwischen Änderungsoptionen festgelegt werden (beispielsweise der gegenseitige Ausschluss zweier Optionen oder dass zwei Optionen voneinander abhängen und somit nur gemeinsam auf ein Prozessmodell angewendet werden können).

Die Anwendung der Optionen, d.h. die eigentliche Variantenbildung, resultiert aus der Auswertung von Variablen in sogenannten Kontextregeln. Beispielsweise könnte eine Kontextregel besagen, dass in Abhängigkeit vom Vorhandensein einer bestimmten medizinischen Abteilung in einem Krankenhaus eine bestimmte Untersuchung in einen Behandlungsprozess aufgenommen wird oder auch nicht.

Beim Provop-Ansatz existiert letztlich also auch für jede Variante ein eigenes Prozessmodell. Im Gegensatz zum konventionellen Multi-Model-Ansatz entstehen diese aber nicht durch einfaches Kopieren und manuelles Editieren des Prozessmodells, sondern werden auf Basis von Kontextregeln und Änderungsoperationen automatisch aus dem Grundmodell generiert, wobei auch Änderungen in diesem Grundmodell an die einzelnen Variantenmodelle weitergegeben werden.

Als letzte speziell auf GPMV ausgerichtete Methode soll der in [DZ11] beschriebene Ansatz vBPMN erwähnt werden. Auch vBPMN ist ein Vertreter des Strukturansatzes. Ausgangspunkt ist dabei ein Referenzprozess, der dann durch die Anwendung sog. Adaptierungsmuster auf Basis des Prozesskontexts verändert wird. Die Autoren weisen selbst darauf hin, dass Sie dem Grundgedanken des Provop-Ansatzes folgen, allerdings setzt vBPMN anders als Provop nicht auf einer proprietären Notation, sondern auf dem BPMN-Standard auf [DZ11].

Der Ansatz lässt bewusst nur Veränderungen in speziell ausgezeichneten Bereichen des Prozessmodells zu, um eine zu komplexe und unkontrollierbare Variantenbildung zu verhindern. Diese Bereiche werden als adaptierbare Segmente bezeichnet [DZK11]. Um die adaptierbaren Segmente zu kennzeichnen, wurden die in BPMN definierten inneren Ereignisse um die Möglichkeit erweitert, Anfangs- und Endmarkierungen zu tragen.

Erreicht eine Prozessinstanz in ihrem Ablauf nun ein adaptierbares Segment, wird dies über das den Beginn des Segments markierende Ereignis signalisiert. Daraufhin werden die Variablen ausgewertet, die die vorliegenden Rahmenbedingungen abbilden. Liegt eine passende Kontextregel vor, kommt es sofort zur Anwendung eines oder mehrerer Adaptierungsmuster und das adaptierbare Segment wird entsprechend angepasst. Die möglichen Adaptierungsmuster sind in einem Musterkatalog erfasst, der in [DZK11] detailliert vorgestellt wird.

Bei vBPMN handelt es sich um einen mächtigen Vertreter des Strukturansatzes. Ein besonderer Vorteil gegenüber ähnlichen Ansätzen ist in der Formalisierung der anwendbaren Muster als parametrisierbare BPMN2-Fragmente zu sehen. Ferner spricht die gute Separierung von Geschäftsregeln für diesen Ansatz. Wegen der Mächtigkeit des Ansatzes sind aber auch komplexe Modellierungskenntnisse erforderlich.

2.4 Deklarative Methoden

Bei den bisher näher vorgestellten Ansätzen handelt es sich um prozedurale Methoden entweder allgemein zur Geschäftsprozessmodellierung oder zur expliziten Modellierung

von GPMV. Deklarative Methoden nehmen aufgrund der inhärenten Variabilität eine Sonderstellung ein. Einen solchen deklarativen Ansatz verfolgt die in [PA06] und [Pe07] vorgestellte Modellierungssprache ConDec.

Die Methode geht von einer „optimistischen“ Betrachtung des Prozesses aus, d.h. der Nutzer wählt während der Abarbeitung der Prozessinstanz zur Laufzeit die Aktivitäten aus, die ihm angemessen erscheinen. Er kann diese in beliebiger Reihenfolge und so oft wie gewünscht ausführen. Um dieses Verhalten zu unterstützen, würde ein sehr einfaches ConDec-Modell genügen, das einfach alle verfügbaren Aktivitäten auflistet. Allerdings gibt es üblicherweise Regeln zu beachten, die bei der Ausführung der Prozessinstanzen erfüllt sein müssen. Diese Regeln oder Einschränkungen können einem ConDec-Modell hinzugefügt werden. Je mehr Regeln hinzukommen, umso strikter ist das Prozessmodell definiert.

Zur Modellierung der Einschränkungen nutzt ConDec Lineare Temporale Logik (LTL). Damit können Abhängigkeiten zwischen Aktivitäten beschrieben werden, beispielsweise dass einer Aktivität A immer irgendwann eine Aktivität B folgen muss. Ferner kann festgelegt werden, wie oft eine Aktivität in einer Prozessinstanz ausgeführt werden kann. Auch verschiedene Beziehungen zwischen Aktivitäten können festgelegt werden. Beispielsweise legt eine „exclusive choice“-Beziehung zwischen zwei Aktivitäten fest, dass genau eine der beiden Aktivitäten auszuführen ist. Eine „not co-existence“-Beziehung stellt hingegen dar, dass die betreffenden Aktivitäten nicht zusammen innerhalb eine Prozessinstanz ausgeführt werden können. ConDec liefert eine grafische Syntax für typische Einschränkungen in einem GPM.

Beim Vergleich von ConDec mit den bereits vorgestellten Ansätzen darf nicht vergessen werden, dass es sich um eine gänzlich andere Sichtweise auf ein GPM handelt. Die nächste auszuführende Aktivität wird durch den Nutzer der Prozessinstanz zur Laufzeit festgelegt. Außerdem wird der Prozesskontext im Modell nicht explizit dargestellt. Diese Faktoren muss der Nutzer bei der Auswahl der nächsten auszuführenden Aktivität selbst berücksichtigen. Somit ist bei dieser Form der deklarativen GPM-Beschreibung im Gegensatz zu den bereits beschriebenen Ansätzen das Wissen über die Rahmenbedingungen des Prozesses nicht im Modell enthalten. Der Fokus liegt darauf, die zeitlich-logischen Beziehungen zwischen Aktivitäten zu beschreiben.

3 Einordnung der bisherigen Forschung

3.1 Charakterisierung des Übersichtsartikels

[AL14a] bietet einen Überblick über die im Kapitel 2 vorgestellten Methoden zur Beherrschbarkeit der GPMV und ordnet diese in die in Abschnitt 2.1 kurz beschriebene Kategorisierung ein. Der Artikel stellt die Anwendung der Ansätze anhand eines selbst kreierten Beispielprozesses detailliert vor und vergleicht Vor- und Nachteile der

Ansätze.

Grundlage des Artikels war eine Literaturrecherche. Als Ausgangspunkt diente eine Suche mit den Schlüsselwörtern *process*, *variability*, *flexibility*, *variants*, *bpm* und *configuration* bei Google Scholar. Außerdem wurde eine ganze Reihe bereits vorab bekannter Artikel im Bereich der GPMV berücksichtigt. Insgesamt gesichtet wurden dabei 64 Artikel zum Thema, von denen letztlich 33 im Detail untersucht wurden und die Grundlage für die Kategorisierung der Ansätze und die Auswahl der repräsentativen Vertreter bildeten.

Ziel des Artikels war es, einen Überblick über vorhandene Methoden zur GPMV zu schaffen. Außerdem konnten aufgrund der detaillierten Anwendung auf einen gemeinsamen Beispielprozess auch Aussagen über die praktische Anwendbarkeit der Ansätze für überschaubare Problemstellungen gewonnen werden.

Einen Überblick über die genannten Merkmale des Übersichtsartikels bietet Tabelle 1.

Charakteristik	Ausprägung
Gegenstand	Methoden zur GPMV
Art des Artikels	Methodenvergleich, Kategorisierung
Angewandte Methoden	Literaturrecherche, Vergleich anhand eines detaillierten durchgängigen Beispiels
Ziel	Übersicht über vorhandene Methoden zur GPMV und deren praktische Anwendbarkeit für überschaubare Problemstellungen

Tab. 1: Charakterisierung des Übersichtsartikels

3.2 Ergebnisse und Thesen

In diesem Abschnitt werden nun einige der Ergebnisse der bisherigen Untersuchung dargestellt. Anschließend stellen die Autoren auf Grundlage der Ergebnisse eine Reihe von Thesen auf, die als Diskussionsgrundlage für die kritische Auseinandersetzung mit dem vorliegenden Artikel dienen sollen. Tabelle 2 bietet einen Überblick über die bisher beschriebenen Methoden sowie deren grundlegenden Ansatz zum Umgang mit Variabilität.

Die Wahl zwischen einem deklarativen und einem prozeduralen Ansatz wird vor allem durch die angestrebte Art der IT-Unterstützung bestimmt: Besteht das Ziel der IT-Unterstützung für das Geschäft darin, strukturierte Aufgaben zum richtigen Zeitpunkt den richtigen Bearbeitern zuzuordnen, bietet sich häufig die Installation eines Workflow-Managementsystems an. Da diese prozedural organisiert sind, ist es sinnvoll, die Prozessvarianten (die letztlich die Konfiguration des Workflow-Managementsystems definieren) auch als prozedurale GPM zu modellieren. Ist das Hauptziel hingegen, wenig strukturierte Prozesse (typisch etwa bei Wissensarbeitern) IT-technisch zu unterstützen, kann der deklarative Ansatz in Verbindung mit Adaptive Case Management-

Werkzeugen die bessere Wahl darstellen. Hier könnte in Zukunft auch der neue OMG-Standard CMMN (*Case Management Model and Notation*) eine Rolle spielen [OMG13], der sich durchaus für die Modellierung von Variabilität anwenden lässt [HKW14].

Mit steigender Anzahl der Varianten bzw. der in den Varianten modellierten Unterschiede bietet der Strukturansatz bessere Möglichkeiten zum Umgang mit der steigenden Komplexität als der Verhaltensansatz (siehe [DRS13]). Dies liegt vor allem daran, dass beim Strukturansatz die Unterschiede in Änderungsoperationen gekapselt bzw. modularisiert werden, die jeweils den Unterschied zwischen zwei konkreten Varianten beschreiben. Eine einzelne stark abweichende Variante mit zahlreichen Unterschieden fällt damit nicht so stark ins Gewicht wie beim Verhaltensansatz, bei dem auch diese stark abweichende Variante in das erstellte Gesamtmodell der Prozessmodellfamilie einbezogen werden müsste und dieses so unverhältnismäßig aufbläht.

Ansatz	Literaturquelle	Grundlegende Modellierung	Variabilitätsansatz
Single-Model	[To13]	prozedural	Verzweigungen in Gesamtmodell mit allen Varianten
Multi-Model	[To13]	prozedural	Separates Modell für jede Variante
C-EPC	[RA07]	prozedural	Verhaltensansatz
Provop	[HBR10]	prozedural	Strukturansatz
vBPMN	[DZ11]	prozedural	Strukturansatz
ConDec	[PA06]	deklarativ	durch deklarative Beschreibung gegeben

Tab. 2: Systematisierung der vorgestellten Ansätze (in Anlehnung an [AL14a])

Dennoch beschränken sich bisherige Untersuchungen zur Anwendbarkeit der GPMV-Ansätze im Wesentlichen auf beispielhafte Problemstellungen mit überschaubarer Komplexität. Eine Anwendung der GPMV-Ansätze in einem realen Modellierungsprojekt jenseits der jeweiligen akademischen Forschungsprojekte der Autoren der Ansätze würde verschiedene Defizite in Bezug auf die praktische Anwendbarkeit der vorhandenen Ansätze bei Problemstellungen mit gestiegener Komplexität besser aufzeigen. Diese liegen vor allem im Bereich der Verbindung und Rückverfolgbarkeit zu fachlichen Anforderungen, der Darstellung der Abhängigkeiten und Wechselwirkungen zwischen Variabilitätspunkten sowie der Nutzung verschiedener Abstraktionsebenen bei der Modellierung.

These 1: Die vorgestellten Methoden bieten im Kern geeignete Ansätze zum Umgang mit GPMV, müssen aber noch in eine geeignete Gesamtvorgehensweise eingebettet werden, um im Kontext realer Modellierungsprojekte anwendbar zu sein.

Die gegenwärtigen GPMV-Methoden konzentrieren sich hauptsächlich auf recht kleinteilige Änderungen meist auf Ebene einzelner Modellelemente. Wie bereits ausgeführt, variieren aber oft größere Fragmente, was eine Modellierungsmethode wirksam unterstützen müsste. Auch treten Variabilitätspunkte nicht unabhängig voneinander auf. Eine Modellierungsmethode muss also nicht nur die Variabilitätspunkte nennen, sondern auch Zusammenhänge zwischen diesen aufzeigen ("requires"- und "excludes"-Beziehungen zwischen Modelländerungen/Features) [Ge03].

In realen Projekten tritt ferner eine große Zahl von Variabilitätspunkten und Prozessvarianten auf. Eine durch die Autoren vorgenommene exemplarische Untersuchung der Variabilität in Prüfungsprozessen deutscher Hochschulen zeigte beispielsweise, dass sich ca. 10% der Modellelemente zwischen den verschiedenen Varianten unterschieden, obwohl diese sogar auf einer gemeinsamen gesetzlichen Grundlage aufgebaut waren [AL14b]. Hierbei kann der Überblick schnell verloren gehen. Umso wichtiger ist eine ganzheitliche Vorgehensweise, die beispielsweise auch die Herkunft von Variabilität im Sinne eines geeigneten Anforderungsmanagements betrachtet und einen sinnvollen Mechanismus zur Rückverfolgung zu den Anforderungen bietet. Dabei geht es vor allem darum, auch darzustellen, *warum* ein bestimmter Ablauf so sein muss, wie er modelliert wurde. Diese Information fehlt in den vorhandenen Ansätzen bislang.

In [Yu97] wird das i*-Framework beschrieben, mit dem in einer frühen Phase des Anforderungserhebungsprozesses dargestellt wird, welche Ziele die Akteure verfolgen, welche alternativen Möglichkeiten zum Erreichen dieser Ziele existieren und wie eine Entscheidung für eine der Möglichkeiten die verschiedenen Ziele positiv oder negativ beeinflusst. Dieser Grundgedanke ließe sich auch auf GPM-Familien übertragen. Es wäre beispielsweise darstellbar, dass Modellvariante A das Ziel „Sicherheit“ positiv beeinflusst, das Ziel „Benutzerfreundlichkeit“ negativ und dass es bei Modellvariante B umgekehrt ist. Diese Bezugnahme auf die Ziele könnte auch die langfristige Wartung einer GPM-Familie unterstützen. Zwar existieren bereits gewisse Mechanismen, um mit der Kleinteiligkeit von Änderungen an Prozessvarianten umzugehen (siehe z.B. der Zusammenschluss einzelner Änderungsoperationen zu gemeinsam anzuwendenden Optionen bei Provop), aber dennoch ist die Übersicht bei einer hinreichend komplexen GPM-Familie mit den bisherigen Methoden langfristig kaum zu sichern. Es ist nur schwer nachzuvollziehen, wie sich Änderungen insbesondere in den gemeinsamen Prozessbestandteilen der Familie auf die einzelnen Varianten auswirken.

Durch eine solche Einbettung in eine umfassendere Vorgehensweise wäre eine Grundlage für die weitere Verbreitung von GPMV-Ansätzen gegeben. Diese könnten für den Bereich der Geschäftsprozessmodellierung einen ähnlichen Stellenwert einnehmen wie Produktfamilienansätze in anderen Bereichen der Wirtschaft und Praxis. Geschäftsprozessmodellfamilien könnten beispielsweise innerhalb weit verzweigter Konzerne oder Behörden ein hohes Maß an Wiederverwendung und damit einhergehende Kostenersparnisse ermöglichen (siehe [Sc14][Wa14]).

These 2: GPMV-Ansätze bieten eine gute Grundlage für die Wiederverwendung von Geschäftsprozessmodellen und damit einhergehende Kostenersparnisse.

Um dieses Ziel zu erreichen, muss neben der Einbettung des Ansatzes in eine ganzheitliche Vorgehensweise aber auch die Methode im Kern weiterentwickelt werden. Sowohl bei prozeduralen als auch bei deklarativen Ansätzen gibt es spezifische Vor- und Nachteile. Deshalb liegt die These nahe, dass eine Methode, die in geeigneter Weise prozedurale und deklarative Ansätze kombiniert, eine vielversprechende Vorgehensweise zur Modellierung von GPMV darstellt.

These 3: Ein bewusst kombinierter prozedural-deklarativer Ansatz könnte neue Möglichkeiten zum geeigneten Umgang mit GPMV darstellen.

Zu dieser These ist zunächst zu untersuchen, inwieweit bestehende Ansätze bereits eine Kombination von prozeduralen und deklarativen Elementen aufweisen. Auch müsste detailliert betrachtet werden, welche Vorteile sich kombinieren lassen und welche Nachteile ausgeglichen werden können.

In Verbindung damit ist stets zu betrachten, inwieweit trotz der Kombination dieser beiden Grundansätze (und einer damit einhergehenden möglicherweise steigenden Komplexität) die Verständlichkeit und praktische Anwendbarkeit für den Nutzer gewahrt bleiben kann. Eine Möglichkeit dazu wäre es, das Prozessmodell getrennt vom Variantenmodell zu betrachten. Ein deklarativ beschriebenes Variantenmodell zeigt dann die möglichen Verbindungen zwischen prozedural beschriebenen Prozessmodellfragmenten auf. Dabei müsste dann auch Wissen über den Prozesskontext einbezogen werden. Genau dieser Punkt ist derzeit beim deklarativen ConDec-Ansatz noch nicht berücksichtigt (siehe Abschnitt 2.4).

In Verbindung damit mag der Einwand berechtigt sein, warum nicht einfach bereits bestehende und praxiserprobte Ansätze für Produktfamilien in den Bereich der GPMV übertragen werden. Dem ist entgegenzuhalten, dass solche Ansätze in der Regel auf statische Strukturmodelle (wie z.B. Klassenmodelle der objektorientierten Softwareentwicklung) abzielen. Hier läuft die Variabilität oft auf das Vorhandensein oder Nichtvorhandensein bestimmter Elemente hinaus (Klassen, Attribute, Methoden etc.), die ggf. noch über bestimmte Abhängigkeiten miteinander verbunden sind (requires- oder excludes-Beziehungen). Im Bereich der GPMV sind jedoch durch den modellierten Kontrollfluss vielschichtigere Abhängigkeiten zwischen einzelnen Prozessteilen zu berücksichtigen. So kann auch eine schnell formulierte und scheinbar kleine Änderung einen weitreichenden Umbau eines Prozessmodells verursachen. Beispielsweise könnte in einem Prozess zum Einholen von Angeboten gefordert werden, standardmäßig zwei Angebote einzuholen. Diese werden dann in zwei parallelen Strängen über mehrere Schritte bearbeitet. Eine Geschäftsregel könnte nun besagen, dass bei Überschreiten einer bestimmten Kostengrenze stattdessen drei Angebote eingeholt werden müssen. Schon diese an sich geringe Veränderung einer Multiplizität führt zum Einfügen eines dritten parallelen Strangs mit allen notwendigen Einzelschritten, der ggf. auch noch Auswirkungen auf spätere Schritte im Gesamtprozess hat. Aus diesen

Betrachtungen ergibt sich die vierte und letzte These in diesem Beitrag.

These 4: Für GPMV geeignete Ansätze unterscheiden sich von allgemeinen strukturbezogenen Variabilitätsansätzen aus dem Bereich der Produktfamilienmodellierung.

4 Diskussion

Aus den vier aufgestellten Thesen und den damit verbundenen Fragestellungen ergibt sich der weitere Forschungsbedarf im Bereich der GPMV.

Ein wichtiger Punkt dabei ist die Anwendung der existierenden GPMV-Methoden in realen Modellierungsprojekten mit entsprechender Komplexität, um bessere Aussagen zur Anwendbarkeit der Methoden zu gewinnen. Insbesondere existierende Defizite der Methoden sollten dabei systematisch erhoben werden. In Verbindung damit könnte bei Anwendung der GPMV-Methoden in realen Projekten aber auch untersucht werden, wie hoch der erzielte Grad an Wiederverwendung tatsächlich aussieht und welche Einspareffekte damit erzielt werden können. Den Autoren sind derzeit keine Untersuchungen dazu im Rahmen von realen Modellierungsprojekten (abgesehen von den Forschungsprojekten der jeweiligen Autoren der Ansätze) bekannt. Im nächsten Schritt müssten nach Ansicht der Autoren auf Basis der erkannten Defizite Weiterentwicklungen in zwei Richtungen vorgenommen werden:

1. Erweiterung der GPMV-Methoden im Kern durch eine Kombination von prozeduralen und deklarativen Elementen unter besonderer Berücksichtigung der Verständlichkeit und Anwendbarkeit der resultierenden Methode.
2. Einbettung der resultierenden Methode in eine ganzheitliche Vorgehensweise.

Diese ganzheitliche Vorgehensweise berücksichtigt u.a. Aspekte des Anforderungsmanagements (ggf. in Kombination mit Zielmodellen), der Wechselwirkungen zwischen Variabilitätspunkten sowie des geeigneten Umgangs mit großen Varianzfragmenten. Dabei sollte auch geprüft werden, inwieweit sich entsprechende Punkte aus Produktfamilienansätzen in die Welt der Geschäftsprozesse übertragen lassen.

Ziel muss es sein, dass durch diese Vorgehensweise die derzeit noch im Wesentlichen im akademischen Bereich verankerten GPMV-Methoden einen weiteren Reifegrad erreichen, der sie auch in einem Praxisumfeld nutzbar macht.

Literaturverzeichnis

[AL14a] Arnold, O., Laue, R.: Überblick über Ansätze zur Modellierung von Variabilität in Geschäftsprozessmodellen. HMD Praxis der Wirtschaftsinformatik, Band 51, Heft 5/2014, S.627-642, 2014.

[AL14b] Arnold, O., Laue, R.: Untersuchungen zur Beherrschbarkeit von Variabilität in Prüfungsprozessmodellen deutscher Hochschulen. In (E. Plödereder u. a. Hrsg.): Lecture Notes in Informatics, Informatik 2014. Gesellschaft für Informatik, Bonn, S. 1135–1147, 2014.

[DRS13] Döhring, M., Reijers, H., Smirnov, S.: Configuration vs. adaptation for business process variant maintenance: An empirical study. Information Systems, 39, S.108–133, 2013.

[DZ11] Döhring, M., Zimmermann, B.: vBPMN: Event-Aware Workflow Variants by Weaving BPMN2 and Business Rules. In (Halpin, T. u.a., Hrsg.): BPMDS 2011 und EMMSAD 2011 Proceedings. S. 332–341, 2011.

[DZK11] Döhring, M., Zimmermann, B., Karg, L.: Flexible workflows at design-and runtime using BPMN2 adaptation patterns. In (Abramowicz, W., Hrsg.): BIS 2011 Proceedings. S. 25–36, 2011.

[Ge03] Geyer, L.: Variabilitätsmanagement in Produktfamilien. Dissertation, Universität Kaiserslautern, 2003.

[HBR09] Hallerbach, A.; Bauer, T.; Reichert, M.: Issues in Modeling Process Variants with Provop. In (Ardagna, D., Hrsg.): Business Process Management Workshops - BPM 2008 International Workshops, Milano 2008. Springer, Berlin Heidelberg, S. 56–67, 2009.

[HBR10] Hallerbach, A.; Bauer, T.; Reichert, M.: Configuration and Management of Process Variants. In Handbook on Business Process Management. Springer, Berlin Heidelberg. S. 237–255, 2010.

[HKW14] Herzberg, N., Kirchner, K., Weske, M.: Modeling and Monitoring Variability in Hospital Treatments: A Scenario using CMMN. In (Fournier, F., Mendling, J., Hrsg): Business Process Management Workshops 2014. Springer Berlin Heidelberg, S. 3-15, 2014.

[KY12] Kumar, A.; Yao, W.: Design and Management of Flexible Process Variants Using Templates and Rules. In Computers in Industry 63(2), S. 112–130, 2012.

[OMG13] OMG, Case Management Model and Notation (CMMN). http://www.omg.org/spec/CMMN/1.0/Beta1/PDF/, Stand: 23.04.2015

[Pa11] Pascalau, E. et. al.: On Maintaining Consistency of Process Model Variants. In (zur Muehlen, M., Su, J., Hrsg.): Business Process Management Workshops - BPM 2010 International Workshops and Education Track, Hoboken 2010. Springer, Berlin Heidelberg 2011, S. 289–300, 2011.

[Pe07] Pesic, M. et. al.: Constraint-based workflow models: Change made easy. In

(Meersman, R. Tari, Z., Hrsg.) Confederated International Conferences CoopIS, DOA, ODBASE, GADA, and IS 2007 Proceedings. Springer Berlin Heidelberg, S. 77–94, 2007.

[PA06] Pesic, M.; van der Aalst, W.M.P.: A declarative approach for flexible business processes management. In BPM'06 Proceedings. S. 169–180, 2006.

[RA07] Rosemann, M.; van der Aalst, W.M.P.: A Configurable Reference Modelling Language. In Information Systems Databases: Their Creation, Management and Utiliza-tion. Elsevier Science Ltd, Oxford, 2007; S. 1–23.

[Sc14] Schumann, R. et.al.: Effective Business process documentation in federal structures. In (Plödereder, E. u.a. Hrsg.): Proceedings of Informatik 2014 (Workshop BPM im Öffentlichen Sektor). Stuttgart, Germany: GI, S. 1043 – 1058, 2014.

[To13] Torres, V. et.al.: A Qualitative Comparison of Approaches Supporting Business Process Variability. In Business Process Management Workshops, BPM 2012 International Workshops, Tallinn, 2012; Springer, Berlin Heidelberg; S. 560–572, 2013.

[Va13] Valenca, G. et.al.: A Systematic Mapping Study on Business Process Variability. In International Journal of Computer Science & Information Technology 5(1), 2013.

[Vo12] Vogelaar, J. et.al.: Comparing Business Processes to Determine the Feasibility of Configurable Models: A Case Study. In Business Process Management Workshops, BPM 2011 International Workshops, Clermont-Ferrand, 2011. Springer, Berlin Heidelberg, S. 50–61, 2012.

[Wa14] Walser, K.: Das BPM-Standardisierungssystem in der Öffentlichen Verwaltung der Schweiz und dessen Theorie-basierte Evaluierung. In (Plödereder, E. u.a. Hrsg.): Proceedings of Informatik 2014 (Workshop BPM im Öffentlichen Sektor). Stuttgart, Germany: GI, S. 1161–1174.

[Yu97] Yu, E.S.K.: Towards modelling and reasoning support for early-phase requirements engineering. In Proceedings of ISRE '97: 3rd IEEE International Symposium on Requirements Engineering, S. 226-235, 1997.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Workshop zur Gesundheitstelematik-infrastruktur sowie zu ethischen, juristischen und sozialen Aspekten informatischer Systeme in Pflege und Gesundheits-versorgung (GTI ELSA)

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Workshop zur Gesundheitstelematikinfrastruktur sowie zu ethischen, juristischen und sozialen Aspekten informatischer Systeme in Pflege und Gesundheitsversorgung (GTI ELSA)

Paul Drews[1], Ali Sunyaev[2] und Karsten Weber[3]

Abstract: Der Workshop GTI ELSA widmet sich aktuellen Herausforderungen und Fragestellungen beim Einsatz von Informations- und Kommunikationstechnik (IKT) im Gesundheitswesen. Mit der Gesundheitstelematikinfrastruktur und altersgerechten Assistenzsystemen werden zwei thematische Schwerpunkte aus diesem Bereich behandelt. In Deutschland wird seit über 10 Jahren die Einführung der Anwendungen der elektronischen Gesundheitskarte (eGK) gemäß § 291a SGB V geplant und vorbereitet. Inzwischen wurde die eGK flächendeckend an die Versicherten der gesetzlichen Krankenversicherung ausgegeben. Ihre Nutzung ist seit Anfang 2015 verpflichtend. Derzeit wird mit dem Aufbau der Gesundheitstelematikinfrastruktur (GTI) der sogenannte Online-Rollout vorbereitet und damit eines der größten Projekte zur IT-gestützten Vernetzung einer Branche vorangetrieben. Spätestens mit der Ausschreibung des Förderschwerpunkts „Altersgerechte Assistenzsysteme“ durch das BMBF im Jahr 2008 wurden der demografische Wandel, dessen gesellschaftliche Auswirkungen und die Beherrschung dieser Wirkungen zu Herausforderungen, denen auch technisch zu begegnen sei. Ambient Assisted Living bzw. altersgerechte Assistenzsysteme sollen, der ursprünglichen Konzeption folgend, insbesondere die Gesundheitsversorgung und Pflege zuhause unterstützen; inzwischen jedoch werden sehr viele verschiedene Systeme unter diesen Bezeichnungen subsumiert. Der Einsatz von IKT im Bereich der Pflege- und Gesundheitsversorgung wirft zahlreiche Fragen auf; die Art der technischen Gestaltung solcher Systeme bringt erhebliche Herausforderungen mit sich, da hier Technik im Umfeld einer besonders vulnerablen Personengruppe eingesetzt wird. Im Rahmen des Workshops werden darüber hinaus ethische, juristische und soziale Aspekte (ELSA) sowie ökonomische Fragen des Einsatzes von IKT im Kontext der Pflege- und Gesundheitsversorgung behandelt.

Organisation: Paul Drews (Leuphana Universität Lüneburg), Ali Sunyaev (Universität zu Köln), Karsten Weber (BTU Cottbus-Senftenberg, OTH Regensburg)

Programmkomitee (ELSA): Elske Ammenwerth (University for Health Sciences, Medical Informatics and Technology, Hall/Tirol), Juliane Eichhorn-Kissel (BTU Cottbus-Senftenberg), Heiner Fangerau (Universität zu Köln), Ursula Hübner (Hochschule Osnabrück), Michael Marschollek (Peter L. Reichertz Institut für Medizinische Informatik, Hannover), Karsten Weber (BTU Cottbus-Senftenberg, OTH Regensburg), Alfred Winter (Universität Leipzig)

[1] Leuphana Universität Lüneburg, IEG, Scharnhorststr. 1, 21335 Lüneburg, paul.drews@leuphana.de
[2] Universität zu Köln, ISQ, Pohligstr. 1, 50969 Köln, sunyaev@wiso.uni-koeln.de
[3] BTU Cottbus-Senftenberg, IST, Postfach 101344, 03013 Cottbus, karsten.weber@b-tu.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Haltung und Übertragung von Patientendaten im Cloud Computing: Anforderungserhebung und prototypische Implementierung eines Verschlüsselungsframeworks

Marc Walterbusch, Tim Hoffmann und Frank Teuteberg[1]

Abstract: Ziel des Beitrags ist es auf Basis eines systematischen Literaturreviews sowie eines Experteninterviews Anforderungen für die Übertragung und Haltung von Patientendaten zu identifizieren. Auf Basis dieser Anforderungen wird prototypisch ein Verschlüsselungsframework implementiert und anhand zweier Experteninterviews konzeptionell sowie bei einem Arzneimittellieferanten mittels zweier Anwendungsbeispiele (Verschlüsselte Backups sowie Arzneimittelbestellungen) technisch evaluiert. Auf Basis des systematischen Literaturreviews lässt sich statuieren, dass eine gründliche Betrachtung der deutschen Gesetzgebung hinsichtlich Sicherheit und Datenschutz in der Literatur bisher ausbleibt. Diese Lücke soll mit diesem Beitrag geschlossen werden.

Keywords: Patientendaten, Übertragung, Cloud Computing, Verschlüsselung, Framework, Literaturreview, Experteninterview, Prototypische Implementierung, Anwendungsfall.

1 Einleitung und Motivation

Cloud Computing Services nehmen einen wichtigen Platz im privaten sowie im Geschäftsleben ein. In diesem Zusammenhang bietet das Cloud Computing auch neue grundlegende Möglichkeiten im Gesundheitssektor [EFZ13]. Hierbei sind verschiedene Szenarien denkbar, wie der Austausch von Patientendaten bei der Überweisung von Patienten zwischen Arztpraxen, die Verwaltung von elektronischen Patientenakten oder das Einsehen aktueller medizinischer Daten auf mobilen Geräten [CH11, Ch12]. Zudem können die vom Cloud Computing bekannten Potentiale wie Kosteneinsparungen, Effizienzvorteile und die Entwicklung neuer Geschäftsmodelle auf den Gesundheitssektor transferiert werden [LSW10, ZL10]. Jedoch verhindern aktuell noch Bedenken bezüglich Sicherheit und Datenschutz die breite Einführung in diesem Bereich [EFZ14]. In diesem Kontext sei auf die regulatorischen Vorschriften des Gesetzgebers hingewiesen, der strikte Vorgaben für die Erhebung, Verarbeitung und Nutzung von Patientendaten macht. Aus diesen regulatorischen Vorschriften sind entsprechend Anforderungen an Cloud Computing Services im Gesundheitssektor abzuleiten. Eine essentielle Maßnahme zur Gewährleistung von Datensicherheit ist die Verschlüsselung von Daten. Hierfür gibt es Ansätze in der Literatur wie z. B. das Trusted Cloud Transfer Protocol (TCTP), welches entwickelt wurde um den Kommunikationskanal zwischen

[1] Universität Osnabrück, Unternehmensrechnung und Wirtschaftsinformatik, Katharinenstr. 1, 49096 Osnabrück, {marc.walterbusch;thoffmann;frank.teuteberg}@uni-osnabrueck.de

User Agent (bspw. Browser) und Cloud zu verschlüsseln [Sl13]. Weiterhin sind in der Literatur Ansätze zu Sicherheitsframeworks zu finden, die Anforderungen an Cloud Computing Systeme im Gesundheitssektor aus anderen Quellen zusammenfassen [EFZ13]. Jedoch gibt es keine Frameworks, die den gesamten Zyklus von Systemen bspw. zum Speichern von Patientendaten zu Backup-Zwecken in der Cloud betrachten. Außerdem bleibt eine gründliche Betrachtung der deutschen Gesetzgebung hinsichtlich Sicherheit und Datenschutz in der Literatur bisher aus. Diese Lücke soll mit diesem Beitrag geschlossen werden. Basierend auf einem Literaturreview, unter Berücksichtigung der deutschen Gesetzgebung sowie im Rahmen eines Experteninterviews mit einem Arzneimittellieferanten wird prototypisch ein Verschlüsselungsframework entwickelt und evaluiert.

2 Literaturreview

2.1 Verwandte Arbeiten

Es wurde ein systematischer Literaturreivew mit dem Suchterm *("Best Practice*" OR encryption OR cryptography OR Verschlüsselung) AND (health OR medical OR Gesundheit*) AND (patient OR "patient data" OR data) AND cloud** durchgeführt [Br09, WW02]. Durchsucht worden sind der Top10-Tier der AIS [As15] sowie die fünf internationalen Wirtschaftsinformatikkonferenzen ECIS, ICIS, AMCIS, HICSS, und PACIS mit dem Ergebnis von 71 Literaturquellen, von denen 23 als relevant eingestuft werden konnten. Aufbauend auf den initialen Ergebnissen wurde eine Rückwärtssuche mit 22 relevanten Ergebnissen sowie eine Vorwärtssuche mit 1 relevanten Artikel durchgeführt [WW02].

Verwandte Arbeiten sind das *Trusted Cloud Transfer Protocol* (TCTP) [Sl13] sowie die Abhandlung zum Thema *Security and Privacy System Requirements for Adopting Cloud Computing in Healthcare Data Sharing Scenarios* [EFZ13]. Das TCTP ist ein Datenübertragungsprotokoll, das hauptsächlich für die Verwendung von Browser- bzw. Hypertext Transfer Protocol (HTTP)-basierenden Software as a Service (SaaS) Cloud Computing Lösungen entwickelt wurde [Sl13]. Slawik et al. [Sl14] zeigen beispielhaft, wie das TCTP in einer medizinischen SaaS Lösung verwendet werden kann, sodass die sichere Übertragung sensibler Daten gewährleistet wird. Im Gegensatz zum TCTP wird in dieser Arbeit ein allein lauffähiges Verschlüsselungsframework entwickelt, das selbstverwaltend ein Verschlüsselungsverfahren anwendet. Weitergehend geht es bei den Anwendungsfällen des Verschlüsselungsframeworks nicht ausschließlich um die Verschlüsselung der Datenübertragung, sondern ebenfalls um die verschlüsselte Datenspeicherung in der Cloud, die das TCTP nicht unterstützt. Bei dem zweitgenanntem Artikel [EFZ13] handelt es sich um eine verwandte Forschungsarbeit, die Sicherheits- und Datenschutzanforderungen an das Cloud Computing im Gesundheitssektor behandelt. Dafür gehen Ermakova et al. [EFZ13] nach dem systematischen Design Science Ansatz vor, wonach zunächst mittels eines

Literaturreviews eine Anforderungserhebung durchgeführt wird. Danach erfolgt die Anwendung der festgestellten Sicherheitsanforderungen mittels Szenarioanalyse. Die von Ermakova et al. [EFZ13] identifizierten Sicherheits- und Datenschutzanforderungen betreffen unmittelbar auch diese Arbeit, in der ein Verschlüsselungsframework für das Cloud Computing im Gesundheitssektor entwickelt werden soll, welches die genannten Anforderungen zum Teil abdecken muss. Hierbei gilt es jedoch zu differenzieren, da nicht alle Sicherheitsproblematiken durch eine Verschlüsslung gelöst werden können. Schließlich kann bspw. das Fehlen eines Notfallzugriffs auf das System oder das Logging des Benutzerverhaltens nicht durch ein Verschlüsselungsframework übernommen werden. Jedoch muss eine Verschlüsselung die Verhinderung eines unautorisierten Zugriffs unterstützen und bei Angriffen oder Datendiebstahl ausreichend stark sein, damit Dritte keinen Zugriff auf den Klartext erlangen. Hierbei grenzt sich die vorliegende Arbeit durch die zusätzliche Identifikation von Regularien und Best Practices ab. Außerdem erfolgt durch die Konsolidierung der Regularien und Best Practices die Ableitung konkreter Anforderungen an ein Verschlüsselungsframework für das Cloud Computing im Gesundheitssektor.

2.2 Best Practices

Auf Basis der Ergebnisse des systematischen Literaturreviews ist eine Konzeptmatrix der Best Practices erstellt worden.[2] Die am häufigsten genannte Best Practice ist die Unterstützung eines adäquaten (i) *Schlüsselmanagements* seitens des Verschlüsselungsframeworks innerhalb der Kommunikation zwischen Sender und Empfänger [AB09, BK14, OS12, Sl14, ST13, ZL10, Ca11, CH11, Ch12, DNP12, DP11, Ja11, Li10, LSW10]. Hierbei wird eine sichere und sinnvolle Methode zum Austausch der Schlüssel zwischen Sender und Empfänger empfohlen. Weiterhin wird eine sichere Verwaltung und Speicherung der Schlüssel als sinnvoll erachtet, die bspw. durch eine Verschlüsselung der Schlüssel erfüllt werden kann [AB09, KL10, Li10]. Das (ii) *hybride Verschlüsselungsverfahren* wurde am zweithäufigsten genannt [AB09, CH11, DNP12, Ja11, Li10, LSW10, LYZ13, Mi13, ST13, We14], wobei diese Best Practice mit dem *Schlüsselmanagement* korrespondiert, da hierbei das asymmetrische Verschlüsselungsverfahren genutzt werden soll, um einen symmetrischen Sitzungsschlüssel zwischen Sender und Empfänger auszutauschen. Des Weiteren empfehlen sieben der analysierten Artikel die Verwendung von verbreiteten und gut getesteten (iii) *Verschlüsselungsstandards* für Verschlüsselungsoperationen im Bereich Cloud Computing im Gesundheitssektor [AB09, BK14, Ca11, DNP12, KL10, Sl14, ST13]. Explizit werden der *Advanced Encryption Standard* (AES) mit einem 256 Bit Schlüssel als symmetrischer Verschlüsselungsalgorithmus [AB09, DNP12, KL10, ST13] sowie der nach den Entwicklern Rivest, Shamir und Adleman benannte asymmetrische Verschlüsselungsalgorithmus *RSA* [AB09, Ch12, DP11, ZL10] vorgeschlagen. Die weiteren identifizierten Best Practices umfassen die (iv) *separate Verschlüsselung* (Einträgen, Dateien und Datensätze nicht in Containern zusammengefasst verschlüsseln,

[2] Dem interessierten Leser wird die entsprechende Konzeptmatrix auf Anfrage gerne zur Verfügung gestellt.

sondern lediglich logisch zusammengehörige Daten – bspw. eine Bestellung bestehend aus Artikel, Preis usw. – zusammen zu verschlüsseln) [AB09, BK14, Ch12, LSW10, LYZ13, Pr06], die (v) *Schlüsselerzeugung und Speicherung beim Client* durchzuführen (Geheimhaltung der Schlüssel nach dem *Need-to-know-Principle*[3] [JP00]; es ist nicht nötig, dass geheime Schlüssel in Kontakt mit der Cloud Computing Umgebung kommen) [AB09, BK14, DP11, Ja11, KL10, KS11], das (vi) *homomorphe Verschlüsselungsverfahren* (ein Verfahren, das speziell für das Cloud Computing entwickelt wurde; dieses macht es möglich, verschlüsselte Daten innerhalb einer Cloud Computing Umgebung zu verarbeiten ohne dabei Zugriff auf den Klartext preiszugeben) [Bl11, Me13, We14], sowie die Verwendung eines (vii) *Tagging Systems* (die verschlüsselten Daten mit Bezeichnern versehen, wodurch diese mit geringem Aufwand identifiziert und verarbeitet werden können) zum leichten Auffinden von verschlüsselten Einträgen in der Cloud [AB09, KL10].

2.3 Regularien

Die Identifikation der Regularien erfolgte durch die im systematischen Literaturreview ermittelte Literatur sowie durch weitere Literatur, die mittels einer offenen Recherche nach Regularien im Cloud Computing via Google Scholar identifiziert worden ist. Da der Gesetzgeber eine Differenzierung vorsieht (vgl. §3 Abs. 9 BDSG) erfolgt die Unterteilung in *patientenbezogene* und *personenbezogene Daten*. Die einzelnen Regularien stammen aus der (Muster-)Berufsordnung für die in Deutschland tätigen Ärztinnen und Ärzte (MBO-Ä), dem Bundesdatenschutzgesetz (BDSG), der Strafprozessordnung (StPO), den Technischen Richtlinien des Bundesamts für Sicherheit in der Informationstechnik (BSI-TR), dem Sozialgesetzbuch (SGB) sowie dem Bürgerlichen Gesetzbuch (BGB).[4]

Hinsichtlich *patientenbezogener Daten* besagt § 9 Abs. 1 MBO-Ä [BK14, DEG11, Du11], dass Ärzte einer Schweigepflicht über sämtliche Informationen ihre Patienten betreffend unterliegen. Dieser Paragraph korrespondiert mit § 203 StGB, der das Strafmaß für die Verletzung von Privatgeheimnissen festlegt, und §§ 630a ff. BGB, wonach der Behandlungsvertrag das Patientengeheimnis schützt [BK14]. Weiterhin verpflichtet § 10 Abs. 2 MBO-Ä den Behandelnden auf Verlangen des Patienten, diesem Einsicht in seine Patientendaten zu gewähren [BK14, Du11, Me13, We10]. Zusätzlich korrespondiert dieses Gesetz mit § 630g BGB [BK14], der die Einsichtnahme in die Patientenakte beschreibt und § 34 BDSG, der dem Betroffenen das Recht auf Auskunftsanspruch über gespeicherte personenbezogene Daten zuspricht. Eine grundsätzliche Aufbewahrungsfrist von patientenbezogenen Daten von zehn Jahren ergibt sich aus § 10 Abs. 3 MBO-Ä [BK14] und § 630f Abs. 3 BGB, sofern keine verlängerten Aufbewahrungsfristen durch andere gesetzliche Vorschriften bestehen [BK14]. Außerdem ist laut § 10 Abs. 5 MBO-Ä eine elektronische externe Speicherung

[3] Das *Need-to-know-Principle* besagt, dass ein Subjekt nur Zugriff auf Informationen haben sollte, wenn diese zur Wahrnehmung der Aufgaben des Subjekts notwendig sind [SS94].
[4] Dem interessierten Leser wird die entsprechende Konzeptmatrix auf Anfrage gerne zur Verfügung gestellt.

von Patientendaten grundsätzlich möglich, jedoch sind hierbei durch besondere Sicherungs- und Schutzmaßnahmen die Veränderung, Vernichtung und unrechtmäßige Verwendung zu verhindern [BK14]. Weiterhin kann hier der § 203 StGB angeführt werden, wodurch der Arzt verpflichtet ist den Zugriff des Dienstleisters auf patientenbezogene Daten zu verhindern. In § 3 Abs. 9 BDSG werden Gesundheitsdaten als eine besondere Art personenbezogener Daten klassifiziert, weshalb es hierfür spezifische Gesetzgebungen gibt [BK14, Du11, We10]. § 97 Abs. 2 StPO schreibt indirekt die getrennte Speicherung von medizinischen Daten und anderen Datenformen vor, damit der Beschlagnahmeschutz gewährleistet ist [BK14]. Als Kryptografische Vorgaben für Projekte der Bundesregierung hat das Bundesamt für Sicherheit in der Informationstechnik (BSI) die Richtlinie BSI-TR-03116 [Bs14] herausgegeben. Der erste Teil BSI-TR-03116-1 dieser Richtlinie legt die im Gesundheitswesen verbindlichen Sicherheitsanforderungen und -vorgaben für den Einsatz kryptographischer Verfahren für die elektronische Gesundheitskarte, den Heilberufsausweis und die technischen Komponenten der Telematikinfrastruktur fest [BK14]. Dabei macht das BSI konkrete Empfehlungen zur Verwendung von Verschlüsselungsalgorithmen, zum Schlüsselvereinbarungsverfahren, zur Datenauthentisierung und Schlüsselerzeugung. Weiterhin werden Empfehlungen bezüglich Instanzauthentisierung, elektronischer Signatur, Verschlüsselung von Dokumenten, Kommunikation und Schlüsselmanagement ausgesprochen. Zudem verlangen das Sozialgesetzbuch (§ 67b SGB X [BK14, DEG11, Du11]) und das Bundesdatenschutzgesetz (§ 4a BDSG) in bestimmten Fällen für die Übermittlung von patientenbezogenen Daten eine Einwilligung des Betroffenen in schriftlicher Form. Durch § 42a BDSG wird eine nicht-öffentliche oder öffentliche Stelle, die besondere Arten personenbezogener Daten (z. B. Patientendaten) speichert, bei unrechtmäßiger Übermittlung an Dritte oder wenn Dritte unrechtmäßig Zugriff auf diese Daten erlangen, verpflichtet, diesen Vorfall unverzüglich den Betroffenen und der Aufsichtsbehörde zu melden [Du11, We10]. Der § 630f BGB macht das Führen einer elektronischen Patientenakte möglich und legt Anforderungen (bspw. Aufzeichnung von Änderungen) und die spezifischen Daten, die gespeichert werden sollen, fest.

Zu den *personenbezogenen Daten* betreffenden Regularien zählt § 11 BDSG, welcher die vertraglichen Rahmenbedingungen für die Erhebung, Verarbeitung und Nutzung personenbezogener Daten einer externen Stelle festlegt [BK14, BS12, DEG11, DRS10, Du11, Me13, MS11, Se13, We10]. Hierbei gilt vor allem zu beachten, dass der Auftraggeber für die Einhaltung der Vorschriften des BDSG verantwortlich ist. Weiterhin legen §§ 4b, c BDSG Vorschriften für die Übermittlung personenbezogener Daten in das Ausland fest [BK14, BS12, DRS10, MS11, We10]. Diese ermöglichen die Übertragung personenbezogener Daten in Mitgliedsstaaten der europäischen Union, Vertragsstaaten des Abkommens über den Europäischen Wirtschaftsraum sowie Organe und Einrichtungen der Europäischen Gemeinschaften. Bei nicht Gewährleistung eines angemessenen Schutzniveaus ist die Übermittlung untersagt, es sei denn eine der in § 4c Abs. 1 BDSG genannten Ausnahmen trifft zu. Außerdem wird durch § 4f BDSG die Bestellung eines Beauftragten für Datenschutz für öffentliche und nicht-öffentliche Stellen, die personenbezogene Daten automatisiert verarbeiten oder bei dessen Erhebung, Verarbeitung und Nutzung mindestens 20 Personen beschäftigt sind,

festgelegt [BK14]. Der § 28 BDSG reguliert die Datenerhebung und -speicherung für eigene Geschäftszwecke [BS12, DEG11, Du11, We10]. Das Erheben, Speichern, Verändern oder Übermitteln personenbezogener Daten zur Erfüllung der Geschäftszwecke ist zulässig. Die Übermittlung oder Nutzung für einen anderen Zweck ist unter den Voraussetzungen aus § 28 Abs. 2 BDSG zulässig. Schließlich macht § 35 BDSG Vorschriften für das Berichtigen, Löschen und Sperren von Daten [BK14]. Personenbezogene Daten sind zu berichtigen, wenn sie unrichtig sind und können u. U. jederzeit gelöscht werden, außer in den Fällen des § 35 Abs. 3.

3 Anforderungen

Aus den Best Practices und Regularien lassen sich argumentativ-deduktiv explizit und implizit Anforderungen ableiten [WH07]. Es werden Anforderungen, die im Allgemeinen die Datenhaltung, -verarbeitung und -übertragung betreffen, sowie Anforderungen, die unmittelbar für ein Verschlüsselungsframework für das Cloud Computing im Gesundheitssektor anwendbar sind, unterschieden.

- Allgemeine Anforderungen an Datenhaltung, -verarbeitung und -übertragung
  - *Verschlüsselte Datenübertragung*: Verschlüsselung der Daten beim Versand; Authentifizierung von Sender und Empfänger; Sicherstellung der Integrität der Daten durch Signaturverfahren; geeignetes Verfahren zum Schlüsselaustausch
  - *Verfügbarkeit*: stetiger Zugriff; zu jeder Zeit bereitstehende Möglichkeit der Datenoperationen Berichtigen, Sperren und Löschen
  - *Aufbewahrungsfrist*: Einhaltung gesetzlicher Aufbewahrungsfristen von Gesundheitsdaten (grundsätzlich zehn Jahre; je nach Art der Daten durch andere Gesetze auch länger)
  - *Verschlüsselte Datenspeicherung*: Wahrung des Patientengeheimnisses und Wahrung der ärztlichen Schweigepflicht durch Verhinderung des Zugriffs Dritter durch eine ausreichend starte Verschlüsselung; Verschlüsselungsoperationen nur lokal durchführen
  - *Datentrennung*: Trennung von medizinischen Daten und anderen Datenformen
- Anforderungen an ein Verschlüsselungsframework für das Cloud Computing im Gesundheitssektor
  - *Schlüsselmanagement*: zur Verschlüsselung verwendete Schlüssel müssen so gespeichert werden, dass ein Zugriff Dritter verhindert wird und diese den verschlüsselten Daten zuzuordnen sind
  - *Verschlüsselungsstandards*: Verwendung gängiger Verschlüsselungs-algorithmen (vgl. BSI TR-03116-1 [Bs14])

- *Separate Verschlüsselung*: keine Containerdateien sondern lediglich logisch zusammengehörende Daten zusammengefasst verschlüsseln
- *Instanzauthentisierung*: Authentisierung des Empfängers gegenüber dem Sender vor der Übermittlung von Daten
- *Datenauthentisierung*: Sicherstellung der Integrität der übermittelten Daten
- *Schlüsselerzeugung*: Sicherstellung, dass Schlüssel ausschließlich clientseitig erzeugt werden; mind. 100 Bit Entropie[5] für die Generierung von Schlüsseln

Damit einerseits weitere Anforderungen für ein Verschlüsselungsframework für das Cloud Computing im Gesundheitssektor identifiziert und in den Abschnitten zuvor eruierte Erkenntnisse verifiziert werden können, wurde ein leitfadengestütztes Experteninterview durchgeführt. Gesprächspartner war eine Expertin, welche in der Stabsstelle Abrechnung eines Dienstleisters im Gesundheitswesen tätig ist, wo sie für die Abrechnung und Abrechnungsschwierigkeiten (bspw. Retaxierung und Regressvermeidung) sowie für datenschutzrechtliche Angelegenheiten (bspw. Weitergabe von Patientendaten an Krankenkassen) zuständig ist. Vorher arbeitete die Expertin als Sachbearbeiterin in der Bundesopiumstelle, wo Sie u. A. die Integration des Betäubungsmittelrezepts in die Gesundheitskarte auf datenschutzrechtliche Problematiken untersuchte. Zusammenfassend ist zu sagen, dass durch das Experteninterview keine neuen Anforderungen für ein Verschlüsselungsframework für das Cloud Computing im Gesundheitssektor identifiziert werden konnten. Da sich jedoch die Aussagen der Expertin mit denen durch die Literaturrecherche identifizierten Best Practices und Regularien decken, konnten diese verifiziert werden.

4 Implementierung

Wesentliches Einsatzgebiet des Frameworks soll die Verschlüsselung jeglicher in Arztpraxen anfallender Daten sein. Hierbei geht es u. a. um Daten der Personalverwaltung, Warenwirtschaft sowie insbesondere um patientenbezogene Bestellungen von Medikamenten. Ziel des Verschlüsselungsframeworks ist es dabei eine hohe Sicherheit zu bieten, generisch einsetzbar zu sein sowie eine gewisse Plattformunabhängigkeit bei gleicher Codebasis zu bieten. Letzteres ergibt sich aus der Tatsache, dass Arztpraxen eine heterogene Betriebssystem- als auch Gerätelandschaft darstellen. Damit eine möglichst große Spanne von Betriebssystemen und Geräten abgedeckt werden kann, wird das Verschlüsselungsframework in der Programmiersprache C# als Portable Class Library (PCL) in der Entwicklungsumgebung Visual Studio 2013 entwickelt. Bei der Entwicklung des Verschlüsselungsframeworks sind die konsolidierten Anforderungen entsprechend berücksichtigt worden. Weiterhin sind diverse Verschlüsselungsalgorithmen hinsichtlich der Verwendung im

[5] *Entropie* beschreibt in der Kryptologie den Informationsgehalt von Daten bei der Erzeugung von Zufallszahlen.

Gesundheitssektor anhand eines Kriterienkatalogs (*Empfehlung des BSI Verschlüsselungsstandard*, *256 Bit Schlüssellänge*, *keine schwachen Schlüssel*, *keine bekannten Angriffe*, *Sicherheits-Performance-Verhältnis*, *freie Verfügbarkeit*), welcher sich z. T. wiederum aus den Anforderungen ergibt, bewertet worden.[6] Dem Ergebnis entsprechend wurde zur symmetrischen Verschlüsselung AES und zur asymmetrischen Verschlüsselung RSA verwendet. Dies ist auch der Grund, warum das Framework dem Anwender keine Auswahl zwischen verschiedenen Algorithmen bietet. Hiermit folgen wir u. a. der Empfehlung der Gesellschaft für Telematikanwendungen der Gesundheitskarte mbH [Ge08]. Im Folgenden soll verkürzt auf die Implementierung von Schlüsselgenerierung, -bereitstellung, -austausch und -speicherung eingegangen werden.

Die Klasse `KeyFactory`, die der Schlüsselgenerierung dient, wurde als statische Klasse implementiert. Die Methode `DeriveAESKeyFromString()` leitet einen für den AES kompatiblen Schlüssel von der übergebenen Zeichenkette mit Hilfe eines übergebenen Salts ab. Weiterhin verfügt diese Methode über zwei weitere optionale Parameter, die die Länge des abgeleiteten Schlüssels in Bytes sowie die Anzahl der Iterationen beim Ableiten des Schlüssels angeben. Dabei repräsentieren die Iterationen die Anzahl der Durchläufe, die beim sogenannten Key Stretching [YY05] durchgeführt werden; eine hohe Anzahl an Iterationen erschwert dabei Brute Force Angriffe [Ap12]. Des Weiteren erfolgt durch die Methode `GenerateRSAKeyPair()` die Generierung eines asymmetrischen Schlüsselpaares, das mit dem RSA Algorithmus kompatibel ist. Hierbei muss als Parameter die Länge des Schlüsselpaares übergeben werden. Als Rückgabewert ergibt sich eine Instanz der Schlüsselklasse `CryptographicKey`, die den generierten privaten und öffentlichen Schlüssel vorhält.

Für das Vorhalten und die Verwaltung einzelner Schlüssel sind die Klassen `AESKeyManager` und `RSAKeyManager` zuständig. Die Klasse `AESKeyManager` dient dazu der Klasse `AESCryption` zur symmetrischen Verschlüsselung auf Verlangen den Schlüssel zum Ver- bzw. Entschlüsseln zu liefern. Für diesen Zweck wurde ein öffentliches Property `Key` implementiert, das den symmetrischen Schlüssel zurückgibt. In der Methode zum Import des Schlüssels `ImportKey()` wird als Parameter der symmetrische Schlüssel als Byte Array übergeben, woraufhin eine Instanz der Schlüsselklasse `CryptographicKey` initialisiert wird. Als weitere Klasse zur Schlüsselbereitstellung wurde die Klasse `RSAKeyManager` implementiert. Diese stellt der Klasse `RSACryption` das Schlüsselpaar zur asymmetrischen Verschlüsselung zur Verfügung; hierfür wird der öffentliche separat vom privaten Schlüssel vorgehalten.

Der Kern der Implementierung ist der Austausch eines Schlüssels zwischen Client und Server, der zum Verschlüsseln sämtlicher Nachrichten einer Sitzung angewendet wird. Um den Ablauf des Handshakeprozesses (siehe *Abb. 1*) in einzelne Schritte aufzuteilen, wurden bei der Implementierung Nachrichtencodes eingeführt, die innerhalb der Klasse `SessionKeyExchange` genutzt werden, um entsprechende Operationen auszuführen.

[6] Die Bewertung ist nicht Teil dieser Publikation. Dem interessierten Leser werden die entsprechenden Informationen auf Anfrage gerne zur Verfügung gestellt.



Abb. 1: Prozess des Schlüsselaustauschs

Die Klasse `EncryptedKeyStore` fungiert als Speicher für symmetrische und asymmetrische Schlüssel. Zudem hält die Klasse den Pfad zu einer XML Datei vor, in der sämtliche Schlüssel chiffriert gespeichert werden. Die Speicherung der Schlüssel zur Laufzeit erfolgt mittels zweier Attribute vom Typ `Dictionary<String, AESKeyManager>` und `Dictionary<String, RSAKeyManager>`. Dies ermöglicht das komfortable Hinzufügen, Lesen und Löschen von Schlüsseln. Der Konstruktor erwartet die Parameter `AESCryption`, mit der die XML Datei, in der die Schlüssel gespeichert werden, verschlüsselt wird, sowie eine Zeichenkette mit dem Pfad zur XML Datei. Der Schlüssel für die Instanz der `AESCryption` sollte von dem Benutzerpasswort abgeleitet werden. Nach der Initialisierung der Klasse `EncryptedKeyStore` muss zunächst die Methode `Load()` aufgerufen oder neue Schlüssel hinzugefügt werden. Weiterhin ist es möglich durch den Aufruf der Methode `Safe()` sämtliche Schlüssel persistent und verschlüsselt in die zuvor übergebene XML Datei zu schreiben. Der Aufbau der XML Datei kann in *Abb. 2* eingesehen werden, wobei übersichtshalber die Schlüssel gekürzt wurden.

```
<Keys>
   <AES>
     <key id="backupkey">/VILoJt034ypkzhQEbsKJ5kikey>
   </AES>
   <RSA>
     <keypair id="rsa_client_1">
       <public>Z02bQP7k+uLv4bFHWBtCJDmb+</public>
       <private>C47p3Maltbcl1uK8SoBseI4UY</private>
     </keypair>
   </RSA>
</Keys>
```

Abb. 2: Struktur der XML Datei zum Speichern von Schlüsseln (gekürzt)

Um der Klasse Schlüssel hinzuzufügen kann die Methode `Add`, von der es zwei Überladungen gibt, die einerseits einen Schlüssel in der Form eines `AESKeyManager` und andererseits in der Form eines `RSAKeyManager` annimmt, verwendet werden. Weiterhin ist das Lesen von Schlüsseln durch die Methoden `GetRSAKeyManager()` und `GetAESKeyManager()` möglich, die jeweils den Identifikator eines Schlüssels als Zeichenkette übergeben bekommen müssen. Schließlich kann durch den Aufruf der Methode `Remove()` mit einer entsprechenden Zeichenkette, die einen Schlüssel identifiziert, ein Schlüssel gelöscht werden, sofern dieser in der Datenstruktur existiert.

5 Konzeptuelle und technische Evaluation

Die konzeptuelle Evaluation des Verschlüsselungsframeworks ist mit Hilfe von zwei leitfadengestützten Experteninterviews durchgeführt worden [Wa06]. Interviewpartner A arbeitet seit drei Jahren als Softwareentwickler mit dem Fokus der Entwicklung von Webseiten mit HTML, CSS und JavaScript sowie der Entwicklung von Backends mit C#, womit auch das Verschlüsselungsframework programmiert worden ist. Interviewpartner B arbeitet seit vier Jahren als Softwareentwickler mit dem Fokus zum einen auf Web Anwendungen, Webseiten und Windows Anwendungen, zum anderen auf mobilen Anwendungen für Windows Phone, iOS und Android. Letztere werden in der Programmiersprache C# entwickelt. Da in diesem Kontext vor allem auch Portable Class Libraries eine Rolle spielen, eignet sich der Experte zur Evaluation des Frameworks durch seine Erfahrung im Umgang mit diesen.[7]

Das Verschlüsselungsframework ist mit den beiden Interviewpartnern A und B hinsichtlich der *Plattformabhängigkeit*, dem zugrundeliegenden *Klassendiagramm* sowie hinsichtlich der *Umsetzung der Implementierung* (u. a. verwendete Bibliotheken, Schlüsselgenerierung, Ver- und Entschlüsselung, Schlüsselaustausch) evaluiert worden. Aufgrund der Implementierung des Frameworks als PLC ist eine universelle Einsetzbarkeit in verschiedenen Softwareprojekten möglich [A, B]. Weiterhin erlaubt dies den Einsatz des Frameworks auf verschiedenen Betriebssystemen und Endgeräten wie Smartphones mit Android, Windows Phone oder iOS sowie auf Desktop-PCs mit MacOS oder Windows. Außerdem ermöglicht die Ver- und Entschlüsselung verschiedener Datentypen und Dateien einen vielseitigen Einsatz für verschiedene kryptografische Aufgaben. Auch ist der Anwender hier nicht eingeschränkt, da ebenfalls sehr generelle Datentypen wie Byte Arrays ver- und entschlüsselt werden können, in die praktisch jeder Datentyp konvertiert werden kann [A]. Des Weiteren wird der Einsatz des Verschlüsselungsframeworks als Middleware innerhalb des Anwendungsfalls der Bestellung von Arzneimitteln als sinnvoll erachtet. Dies wird damit begründet, dass der Anwender des Frameworks nicht selbst direkt die Funktionaltäten zur Ver- und Entschlüsselung aufrufen muss, sondern dieser wie gewohnt mit den Operationen der Web API arbeiten kann [A; B]. Einige Verbesserungspotentiale sind bereits im Prototypen umgesetzt worden. Weitere Anpassungen werden in der nächsten Iteration der Entwicklung des finalen Verschlüsselungsframeworks umgesetzt.

Die technische Evaluation erfolgte mittels der Nutzung des Verschlüsselungsframeworks bei einem Arzneimittellieferanten in zwei exemplarischen aber realen Anwendungsfällen. Der Arzneimittellieferant gibt seinen Kunden (in diesem Fall Arztpraxen) ein speziell entwickeltes Praxissystem an die Hand. Dieses Praxissystem übernimmt u. a. die Warenwirtschaft und Bedarfsplanung, sodass hiermit elektronische Bestellungen möglich sind. Später soll diese Software noch um weitere Module erweitert werden, sodass sämtliche Aufgabenbereiche einer Arztpraxis abgedeckt werden. Weiterhin soll die Software später serverseitig an eine Cloud angebunden werden,

[7] Im Folgenden wird *Interviewpartner A* bzw. *Experte A* mit *A* abgekürzt; *Interviewpartner B* analog.

weshalb diese zukünftigen Pläne bei den folgenden Ausführungen bereits berücksichtigt werden. Arzneimittelbestellungen sind oftmals patientenbezogen und enthalten damit personenbezogene Daten. Deshalb müssen diese speziell gesichert bzw. Ende-zu-Ende verschlüsselt werden.

Szenario 1: Verschlüsselte Backups

Im Folgenden soll der Einsatz des entwickelten Verschlüsselungsframeworks beim Anlegen und Wiederherstellen von verschlüsselten Backups beschrieben werden.[8] Zu Beginn des Backups muss der entsprechende Schlüssel geladen werden, welcher bei der Installation der Software angelegt wurde. Zum Verschlüsseln der Backups wird der symmetrische Verschlüsselungsalgorithmus AES verwendet, da dieser bei einer Schlüssellänge von 256 Bit als sehr sicher gilt und für die Ver- bzw. Entschlüsselung großer Datenmengen geeignet ist. Für die Verschlüsselung von Dateien, die auf dem System gespeichert werden, nehmen die entsprechenden Methoden jeweils als Parameter den Pfad zur Datei, die verschlüsselt werden soll, und den Zielpfad der verschlüsselten Datei. Hierbei gilt zu beachten, dass die vorhandene Ordnerstruktur beibehalten wird und zu entscheiden, ob Dateinamen verschlüsselt werden sollen oder im Klartext gesichert werden. Das Wiederherstellen eines Backups erfolgt entsprechend auf dem entgegengesetzten Weg mittels der entsprechenden Methode. Sollen Softwarekonfigurationen und Benutzereinstellungen gesichert werden, handelt es sich oftmals um Datenbankdatensätze die per Datenbankabfrage gelesen werden. Sofern ein Backup dieser Konfigurationsdaten nicht mittels Datenbankauszug in eine Datei erfolgt, gestaltet sich die Verwendung der vorherigen Methoden als eher unhandlich, da hierfür die zusätzliche Zwischenspeicherung der aus der Datenbank extrahierten Daten in einer Datei nötig ist. Daher empfiehlt sich für diesen Zweck die Verwendung von Methoden, die einen Datenstrom, der die entsprechenden Konfigurationsdaten aus der Datenbank enthält, ver- bzw. entschlüsseln. Die Speicherung der verschlüsselten Konfigurationsdaten kann schließlich in einer Textdatei oder ähnlichem auf dem externen Medium erfolgen.

Szenario 2: Arzneimittelbestellungen

In diesem Abschnitt wird die Anwendung des Verschlüsselungsframeworks bei der Bestellung von Arzneimitteln erläutert. Hierfür soll ein über HTTP erreichbarer Webserver verwendet werden, der Bestellungen entgegennimmt. Das Aufgeben von Bestellungen soll mit einem einfachen HTTP-Client möglich sein. Der detaillierte technische Aufbau ist in *Abb. 3* ersichtlich.

[8] Da zunächst die Daten des Praxissystems für Arztpraxen auf externen Festplatten gesichert werden und erst zu einem späteren Zeitpunkt in der Software das Speichern von Backups in der Cloud ermöglicht werden soll, wird im Folgenden die Sicherung auf externen Medien betrachtet. Jedoch sind die Unterschiede zwischen dem Speichern in der Cloud oder auf externen Medien vom Standpunkt der Verschlüsselung betrachtet marginal, weshalb die folgenden Beschreibungen ebenfalls auf das Sichern von Backups in der Cloud anwendbar sind.

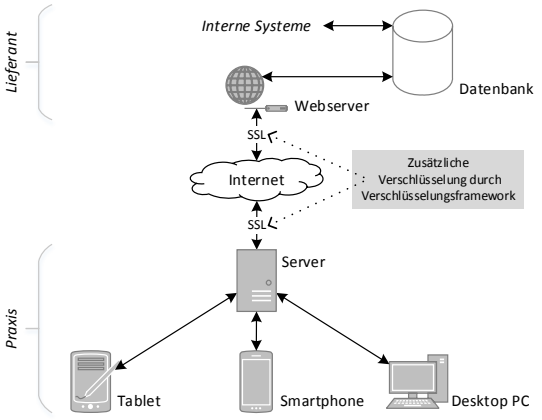


Abb. 3: Technischer Aufbau der Kommunikation

Zur Illustration eines solchen Systems wurde ein ASP .NET MVC Web API [Mi14] Projekt erstellt, das die serverseitige Annahme von Bestellungen übernimmt. Dies ermöglicht das Versenden von Objekten als JavaScript Object Notation (JSON), welche vor dem Versenden mittels HTTP serialisiert werden, sodass diese nach dem Empfangen von der ASP .NET Web API deserialisiert werden und dem Entwickler ohne Zutun als Instanzen zur Verfügung stehen. Sendet also ein Client bspw. eine Instanz der Klasse Bestellung an den Server, so muss der Entwickler keine Zeichenkette parsen sondern erhält eine Instanz der Klasse mit denselben Attributen, wie sie vom Client versendet wurde. Um das Verschlüsselungsframework möglichst anwenderfreundlich einzubetten, wird die Verschlüsselung in der Web API als Middleware implementiert. Hierfür bieten sich sogenannte Delegating Handler an, die in den Kommunikationskanal eingefügt werden und das Lesen und Manipulieren von HTTP-Anfragen sowie Antworten ermöglichen. Innerhalb der Methode findet zunächst die Identifikation des Clients anhand der Session ID statt, woraufhin geprüft wird, ob mit diesem Client bereits der Handshake abgeschlossen ist. Sofern dieses der Fall ist, erfolgt die Entschlüsselung der Anfrage mittels des Sitzungsschlüssels und die Weitergabe zur Verarbeitung. Nachdem die Anfrage verarbeitet und eine Antwort zurückgegeben wurde, wird diese mit dem Sitzungsschlüssel verschlüsselt sowie zum Versand an den Client zurückgegeben. Ist der Handshake bzw. Austausch des Sitzungsschlüssels mit diesem Client noch nicht abgeschlossen, wird, sofern es sich bei der HTTP-Anfrage um eine Handshake Nachricht handelt, der nächste Schritt im Handshakeprozess ausgeführt und die entsprechende HTTP-Antwort zurückgegeben. Dies geschieht solange bis der Handshake erfolgreich war oder fehlgeschlagen ist. Auf Seiten des Clients wird eine Verbindung zu der ASP .NET Web API hergestellt. Weiterhin erfolgt die Verschlüsselung der HTTP-Anfrage, die gesendet werden soll, und die Entschlüsselung der HTTP-Antwort mittels des Sitzungsschlüssels. Da der ASP .NET Web API Server eine Initiierung des Handshakes seitens des Clients erwartet, wurde eine entsprechende Klasse implementiert, die den Handshake startet und den gesamten Handshake durchführt, bis dieser erfolgreich war oder fehlschlägt.

6 Zusammenfassung und Ausblick

Auf Basis eines Literaturreviews und den daraus resultierenden Best Practices und Regularien sind Anforderungen an ein Verschlüsselungsframework im Gesundheitssektor abgeleitet worden. Unter Anwendung dieser Anforderungen ist ein solches Verschlüsselungsframework prototypisch in C# als PLC implementiert worden, welches auf den meisten gängigen Computer- und Handheldplattformen verwendbar ist. Mit Hilfe von Experteninterviews ist das Verschlüsselungsframework konzeptionell sowie in zwei Anwendungsfällen technisch evaluiert worden.

Mit Blick auf Limitationen sei angemerkt, dass der systematische Literaturreview primär auf Beiträge aus der Wirtschaftsinformatik abgezielt hat, korrespondierende Beiträge aus dem Bereich der Medizininformatik/eHealth sind nicht aufgegriffen worden. Weitergehen sei erwähnt, dass jede Verschlüsselung mit entsprechender Rechenleistung zu entziffern ist. Allerdings bietet die Wahl von hohen Schlüssellängen eine hohe Sicherheit. Hinsichtlich der Implementierung ist das Fehlen des Ablaufs eines Sitzungsschlüssels nach einer bestimmten Zeit als kritisch zu betrachten. Hier sollte der Schlüssel seitens des Webservers mit einem Zeitstempel versehen werden, sodass nach einer bestimmten Zeit ein neuer Sitzungsschlüssel ausgetauscht wird. Zudem wurde u. a. die Anforderung der *Datenauthentisierung* an ein Verschlüsselungsframework gestellt, diese allerdings nicht explizit umgesetzt. Hierbei handelt es sich jedoch um kein schwerwiegendes Problem, da bei einer Manipulation der verschlüsselten Daten die Entschlüsselung schlichtweg fehlschlägt und die Datenübertragung abgebrochen wird. Des Weiteren empfiehlt es sich das gesamte Verschlüsselungsframework sowie den vollständig implementierten Anwendungsfall gängiger kryptografischer Angriffe zu unterziehen und zum anderen zu versuchen bspw. als Angreifer den Handshakeprozess zu manipulieren, damit ggf. Schwächen oder Fehler in der Implementierung gefunden und daraufhin behoben werden können.

Bezüglich zukünftigen Forschungsbereichen sollten Vorschläge für Standards bei der Übertragung und Speicherung von Patientendaten (in der Cloud) entwickelt werden, um so ein einheitliches Sicherheitsniveaus zu erreichen. Diese Arbeit hat dafür erste Schritte aufgezeigt, jedoch sind weitere Arbeiten hinsichtlich Übertragungsprotokolle sowie der Integration der Verschlüsselung in die Datenübertragung nötig. Ein weiterer Bereich ist das Teilen von Patientendaten mit Dritten (bspw. bei der Überweisung von Patienten).

Danksagung

Die Autoren danken sowohl der LM IT Services AG, den Gutachtern sowie den Experten, die für die Interviews zur Verfügung standen. Darüber hinaus gilt unser Dank explizit Frau Linster-Hoffmann für ihre Unterstützung. Diese Arbeit ist Teil des Projekts „Nachhaltiger Konsum von Informations- und Kommunikationstechnologie in der digitalen Gesellschaft - Dialog und Transformation durch offene Innovation“. Das

Projekt wird vom Ministerium für Wissenschaft und Kultur des Landes Niedersachsen und der VolkswagenStiftung aus Landesmitteln des Niedersächsischen Vorab gefördert (Projektnummer VWZN3037).

Literaturverzeichnis

[AB09] Archer, J.; Boehm, A.: Security guidance for critical areas of focus in cloud computing. Cloud Security Alliance S. 1–176, 2009.

[Ap12] Apostol, K.: Brute-force Attack. SaluPress, 2012.

[As15] Association for Information Systems, MIS Journal Rankings, http://aisnet.org/general/custom.asp?page=JournalRankings, Stand: 18.01.2015.

[BK14] Bundesärztekammer; Kassenärztliche Bundesvereinigung: Empfehlungen zur ärztlichen Schweigepflicht , Datenschutz und Datenverarbeitung in der Arztpraxis. Deutsches Ärzteblatt 21, S. 963–972, 2014.

[Bl11] Blumenthal, M.: Is Security Lost in the Clouds? Communications and Strategies 2011.

[Br09] vom Brocke, J.; Simons, A.; Niehaves, B.; Riemer, K.; Plattfaut, R.; Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. 17th European Conference on Information Systems. Verona, Italy, 2009.

[BS12] Borges, G.; Schwenk, J.: Daten- und Identitätsschutz in Cloud Computing, E-Government und E-Commerce. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[Bs14] BSI, BSI-TR-03116, https://www.bsi.bund.de/DE/Publikationen/TechnischeRichtlinien/tr03116/index_htm.html, Stand: 21.10.2014.

[Ca11] Carroll, M.: Secure cloud computing: Benefits, risks and controls. Information Security South Africa 2011.

[CH11] Chen, L.; Hoang, D.B.: Novel Data Protection Model in Healthcare Cloud. 2011 IEEE International Conference on High Performance Computing and Communications. Ieee, S. 550–555, 2011.

[Ch12] Chen, T.-S.; Liu, C.-H.; Chen, T.-L.; Chen, C.-S.; Bau, J.-G.; Lin, T.-C.: Secure Dynamic access control scheme of PHR in cloud computing. Journal of medical systems 36/6, S. 4005–20, Dec. 2012.

[DEG11] Duisberg, A.; Eckhardt, J.; Grudzien, W.: Rechtliche Anforderungen an Cloud Computing – Sichere Cloud Dienste. IT-Gipfel 2011 - Rechtliche Anforderungen an Cloud Computing S. 1–67, 2011.

[DNP12] Deng, M.; Nalin, M.; Petković, M.: Towards trustworthy health platform cloud. Secure Data Management S. 162–175, 2012.

[DP11] Deng, M.; Petkovic, M.: A Home Healthcare System in the Cloud--Addressing Security and Privacy Challenges. 2011 IEEE International Conference on Information Systems, Cloud Computing (CLOUD). S. 1–8, 2011.

[DRS10] Doelitzscher, F.; Reich, C.; Sulistio, A.: Designing Cloud Services Adhering to Government Privacy Laws. International Conference on Computer and Information Technology Cit, S. 930–935, Jun. 2010.

[Du11] Duisberg, A.: Gelöste und ungelöste Rechtsfragen im IT-Outsourcing und Cloud Computing. In (Picot, A., Götz, T., and Hertz, U., Hrsg.): Trust in IT. Springer Berlin Heidelberg, Berlin, Heidelberg, S. 49–70, 2011.

[EFZ13] Ermakova, T.; Fabian, B.; Zarnekow, R.: Security and Privacy System Requirements for Adopting Cloud Computing in Healthcare Data Sharing Scenarios. Americas Conference on Information Systems. S. 1–9, 2013.

[EFZ14] Ermakova, T.; Fabian, B.; Zarnekow, R.: ACCEPTANCE OF HEALTH CLOUDS-A PRIVACY CALCULUS PERSPECTIVE. European Conference on Information Systems. S. 1–13, 2014.

[Ge08] gematik, Verwendung kryptographischer Algorithmen in der Telematikinfrastruktur, https://www.gematik.de/cms/media/dokumente/release_0_5_3/release_0_5_3_datensc hutz/gematik_GA_Spezifikation_Kryptographischer_Algorithmen_V1_3_0.pdf, Stand: 03.03.2015.

[Ja11] Jansen, W. a: Cloud Hooks: Security and Privacy Issues in Cloud Computing. 2011 44th Hawaii International Conference on System Sciences. Ieee, S. 1–10, 2011.

[JP00] Janczewski, L.J.; Portougal, V.: "Need-to-know" principle and fuzzy security clearances modelling. Information Management & Computer Security 8/5, S. 210–217, 2000.

[KL10] Kamara, S.; Lauter, K.: Cryptographic Cloud Storage. Proceedings of the 1st Workshop on RealLife Cryptographic Protocols and Standardization S. 1–14, 2010.

[KS11] Kaletsch, A.; Sunyaev, A.: Privacy engineering: personal health records in cloud computing environments. International Conference on Information Systems. S. 1–11, 2011.

[Li10] Li, M.; Yu, S.; Ren, K.; Lou, W.: Securing personal health records in cloud computing: Patient-centric and fine-grained data access control in multi-owner settings. Security and Privacy in Communication Networks S. 89–106, 2010.

[LSW10] Löhr, H.; Sadeghi, A.-R.; Winandy, M.: Securing the e-health cloud. Proceedings of the ACM International Conference on Health informatics - IHI '10. ACM Press, New York, New York, USA, 2010.

[LYZ13] Li, M.; Yu, S.; Zheng, Y.: Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. IEEE Transactions on Parallel and Distributed Systems 24/1, S. 131 – 143, 2013.

[Me13] de Meer, H.; Diener, M.; Herkenhöner, R.; Kucera, M.; Niedermeier, M.; Reisser, A.; Schryen, G.; Vetter, M.; Waas, T.; Yasasin, E.: Sicherheitsherausforderungen in hochverteilten Systemen. PIK - Praxis der Informationsverarbeitung und Kommunikation 36/3, Jan. 2013.

[Mi13] Mitchell, S.; Ridley, S.; Tharenos, C.; Varshney, U.: Investigating Privacy and Security Challenges of mHealth Applications. Americas Conference on Information Systems. S. 1–9, 2013.

[Mi14] Microsoft Corporation, ASP.NET Web API, http://www.asp.net/web-api, Stand: 03.12.2014.

[MS11] Marnau, N.; Schlehahn, E.: Cloud computing und safe harbor. Datenschutz und Datensicherheit - DuD S. 311–316, 2011.

[OS12] Oetzel, M.; Spiekermann, S.: Privacy-by-design through systematic privacy impact assessment-a design science approach. European Conference on Information Systems. 2012.

[Pr06] Pratt, W.; Unruh, K.; Civan, A.; Skeels, M.: Personal health information management. Communications of the ACM 49/1, S. 51, Jan. 2006.

[Se13] Selzer, A.: Die Kontrollpflicht nach § 11 Abs. 2 Satz 4 BDSG im Zeitalter des Cloud Computing. Datenschutz und Datensicherheit - DuD 37/4, S. 215–219, 2013.

[Sl13] Slawik, M.: The Trusted Cloud Transfer Protocol. 2013 IEEE International Conference on Cloud Computing Technology and Science. Ieee, S. 203–208, 2013.

[Sl14] Slawik, M.; Ermakova, T.; Repschläger, J.; Küpper, A.: Securing Medical SaaS Solutions Using a Novel End-To-End Encryption Protocol. European Conference on Information Systems. S. 1–9, 2014.

[SS94] Sandhu, R.S.; Samarati, P.: Access control: principle and practice. IEEE Communications Magazine 32/9, S. 40–48, 1994.

[ST13] Stark, L.; Tierney, M.: Lockbox: mobility, privacy and values in cloud storage. Ethics and Information Technology 16/1, S. 1–13, Oct. 2013.

[Wa06] Walsham, G.: Doing Interpretive Research. European Journal of Information Systems 15/3, S. 320–330, 2006.

[We10] Weichert, T.: Cloud Computing und Datenschutz. Datenschutz und Datensicherheit - DuD 34/10, S. 679–687, Oct. 2010.

[We14] Wenge, O.; Lampe, U.; Müller, A.; Schaarschmidt, R.: Data Privacy in Cloud Computing–An Empirical Study in the Financial Industry. Americas Conference on Information Systems. S. 1–10, 2014.

[WH07] Wilde, T.; Hess, T.: Forschungsmethoden der Wirtschaftsinformatik: Eine empirische Untersuchung. Wirtschaftsinformatik 49/4, S. 280–287, 2007.

[WW02] Webster, J.; Watson, R.T.: Analyzing the past to prepare for the future: Writing a Literature Review. MIS Quarterly 26/2, S. 13–23, 2002.

[YY05] Yao, F.F.; Yin, Y.L.: Design and analysis of password-based key derivation functions. IEEE Transactions on Information Theory 51/9, S. 3292–3297, Sep. 2005.

[ZL10] Zhang, R.; Liu, L.: Security Models and Requirements for Healthcare Application Clouds. International Conference on Cloud Computing S. 268–275, Jul. 2010.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

How Older People Matter – Nutzer_innenbeteiligung in AAL-Projekten

Cordula Endter[1]

Abstract: Bei der Bewältigung der sogenannten demographischen Herausforderungen wird zunehmend auf die Entwicklung und Anwendung alters- bzw. generationengerechter Technologien gesetzt. Als intelligente Mensch-Maschine-Systeme sollen sie in die Lebenswelt älterer Menschen integriert werden, um diese bei der selbständigen Bewältigung alltäglicher Handlungen zu unterstützen. Neben der technisch anspruchsvollen Entwicklung dieser assistiven Systeme, die im europäischen Kontext unter den Begriff Ambient Assistive Living (AAL) subsummiert werden, stellt die Beteiligung von Nutzer_innen am Entwicklungsprozess eine weitere Herausforderung für die Projekte dar. Auf diese Herausforderung reagieren diese dementsprechend mit ganz unterschiedlichen Praktiken der Beteiligung. Anhand von drei AAL-Projekten möchte ich aufzeigen, wie unterschiedlich Nutzer_innenbeteiligung in technischen Innovationsprozessen aussehen kann und wie ältere Testpersonen in diesen Praktiken überhaupt erst zu Nutzer_innen altersgerechter Technik gemacht werden.

Keywords: Ambient Assisted Living, Ethnographie, Nutzer_innenbeteiligung, Age Skript

1 Altersgerechte Technik und das Leitbild von Selbstbestimmung im Alter

In einer Informationsgesellschaft, die geprägt ist von Technologie und Geschwindigkeit, in der technische Innovationen alltägliche Routinen in kürzester Zeit verändern, verwerfen oder neu erfinden, kann sich auch das Alter weder kulturell noch sozial dieser Temporalisierung und Informatisierung entziehen. So sind es auch diese beiden Bereiche – Information und Kommunikation –, welche bei der Entwicklung und öffentlichen Verhandlung von technologischen Assistenzsystemen vornehmlich aufgegriffen werden.

Information und Kommunikation werden neben Sicherheit und Gesundheit als natürliche Bedürfnisse einer wachsenden Zahl älterer Menschen dargestellt und damit der Ausbau dieser Technologien für diese Nutzer_innengruppe mit deren Wunsch nach Teilhabe begründet. Sowohl im Kontext von Information und Kommunikation als auch von Sicherheit und Gesundheit wird dabei verstärkt auf den Einsatz technischer Systeme gesetzt. Dahinter steht die Idee mittels moderner Informations- und Kommunikationstechnologien den Herausforderungen einer alternden Gesellschaft begegnen zu können. Zu diesen Herausforderungen zählen die Sicherstellung von

[1] Kollegiatin der Forschungsinitiative „Alter(n) als kulturelle Konzeption und Praxis“, Heinrich-Heine Universität Düsseldorf, Gebäude 23.21, Raum 46C, Universitätsstr. 1, 40225 Düsseldorf, cordulaendter@gmail.com.

Pflege- und Gesundheitsdienstleistungen als auch Fragen der Teilhabe und gesellschaftlichen Integration beispielsweise im Kontext von Mobilität und Wohnen.

Eine Strategie, auf welche die deutsche Bundesregierung, vertreten durch das *Bundesministerium für Bildung und Forschung* (BMBF), zur Bewältigung dieser demographischen Herausforderungen setzt, ist die Unterstützung durch assistive intelligente Technologien.[2] Seit 2008 fördert sie dazu technische und soziale Innovationen im Bereich umgebungsunterstützender Assistenz/Ambient Assisted Living (AAL), wobei der Schwerpunkt eindeutig auf technischen Innovationen liegt, überblickt man die durch das BMBF getätigten Bekanntmachungen.

AAL selbst muss dabei als ein Sammelbegriff einerseits für ganz unterschiedliche Vorstellungen von altersgerechter Technik und anderseits von Altern selbst verstanden werden, die sich nicht nur in den Texten der Bekanntmachung wiederspiegeln, sondern auch in den Projekten, welche gefördert werden. Während die *VDI/VDE Information und Technik GmbH* (VDI/VDE-IT) als Projektträger des BMBF unter AAL „Konzepte, Dienstleistungen und Produkte, die neue Technologien und soziales Umfeld miteinander verbinden, um die Lebensqualität für Menschen in allen Lebensabschnitten zu erhöhen." (VDI/VDE-IT)[3] versteht, fasst sich das BMBF mit seiner Formulierung „Altersgerechte[r] Assistenzsysteme für ein gesundes und unabhängiges Leben" (BMBF)[4] deutlich kürzer und lässt offen inwieweit technologische und soziale Innovationen hier mitgedacht werden sollen. Auffällig ist jedoch, dass neben der expliziten Unterscheidung zwischen technischen und sozialen Innovationen, auch andere potentielle Nutzer_innengruppen, wie zum Beispiel Menschen mit einem Handicap oder jüngere Pflegebedürftige nicht angesprochen werden.

Diese Fokussierung auf Menschen in einem höheren Lebensalter als Nutzer_innen und deren Lebenswelt als Anwendungskontext irritiert, wäre doch eine breitere Nutzer_innengruppe technisch nicht nur naheliegend, sondern auch effizient. So liegen gerade im Bereich der Disability Studies langjährige Erfahrungen in Entwicklung und Anwendung assistiver Technologien vor. Allein die *Fraunhofer-Allianz Ambient Assisted Living* berücksichtigt weitere Nutzer_innengruppen in ihrer Definition von AAL als „intelligente Umgebungen, die sich selbstständig, proaktiv und situationsspezifisch den Bedürfnissen und Zielen des Benutzers anpassen, um ihn im täglichen Leben zu unterstützen [und die] insbesondere auch älteren, behinderten und pflegebedürftigen Menschen ermöglichen, selbstbestimmt in einer privaten Umgebung zu leben" (Fraunhofer-Allianz Ambient Assisted Living)[5].

[2] Der Förderschwerpunkt ist Bestandteil der Forschungsagenda „Das Alter hat Zukunft" und der Demografiestrategie „Jedes Alter zählt" (siehe dazu http://www.das-alter-hat-zukunft.de/de , letzter Zugriff: 30.06.2014; http://www.bundesregierung.de/Webs/Breg/DE/Themen/Demografiestrategie/_node.html , letzter Zugriff: 30.07.2014)

[3] Siehe http://www.vdivde-it.de/projekte/altersgerechte-assistenzsysteme-fuer-ein-gesundes-und-unabhaengiges-leben (letzter Zugriff: 30.06.2014)

[4] Siehe http://www.mtidw.de/ueberblick-bekanntmachungen/selbstbestimmt-leben (letzter Zugriff: 30.06.2014)

[5] Siehe http://www.fraunhofer.de/de/institute-einrichtungen/verbuende-

Was stattdessen wiederholt wird, ist das Ziel, der wachsenden Mehrheit der älteren Bevölkerung ein selbstbestimmtes Leben im eigenen Zuhause, optimale Gesundheitsversorgung auch fernab medizinischer Institutionen und soziale Teilhabe mittels AAL-Technologien zu ermöglichen. Dabei sollen die Geräte alltägliche Abläufe vereinfachen, Orientierungs- und Unterstützungsangebote bereitstellen und Verhaltensweisen, Gewohnheiten und Handlungen aktiv erkennen, erlernen und unterstützen beispielsweise durch Navigationssysteme, internetbasierte Trainingsprogramme oder selbstlernende Kommunikationssysteme [LI11]. Dazu müssen AAL-Technologien situationsabhängig, nutzerzentriert, adaptiv und unaufdringlich unterstützen. Im Zentrum stehen die Bedürfnisse der Nutzer_innen. Aus ingenieurwissenschaftlicher Sicht stellt sich hierbei die Frage, wie Selbstbestimmung in smarte Technologien übersetzt werden kann, damit ältere Menschen als zukünftige Nutzer_innen dieser smarten Technologien diese selbständig, erfolgreich und zufriedenstellend nutzen können. Diese Frage berührt damit Aspekte der alters- bzw. generationengerechten Usability [Cz97; CL07; FR97; Fi09]. Dazu zählen neben Aspekten nutzerfreundlicher Gestaltung und Usability die Berücksichtigung von Kompetenzen und Ressourcen, Bedürfnissen und Anforderungen [KC95; KMM99; Ha00; Ro01; Me11].

Human Factors und Usability Studies beantworten diese Fragen vornehmlich aus einer ergonomischen Perspektive [FR97; DAP07; SSW10] und weisen darauf hin, dass die Bereitschaft zur Nutzung neuer Technologien vorhanden ist [CS98; Mi10] auch wenn diese maßgeblich durch die bisherigen Erfahrungen im Umgang mit Computern oder ähnlichen Geräten, verfügbarem Training und soziale Unterstützung bestimmt sind [Cz97].[6]

Offen bleibt, inwieweit die Einführung technischer Unterstützungssysteme den Einfluss von Informations- und Kommunikationstechnologien auf die Lebenswelt älterer Menschen ausweitet. Diese Ausweitung kann den Druck auf das alternde Subjekt, sich diesen Technologien gegenüber anpassen bzw. sie sich aneignen zu müssen, verschärfen. Damit würde die Lebensphase „Alter“ unter einen neuen Optimierungsdruck geraten, der sich in der öffentlichen Rede vom erfolgreichen oder gelingenden Alter wiederspiegelt [DL12; De14; DL09; Sc00].

Aus diesem Grund möchte ich hier eine Definition vorschlagen, die AAL als eine soziale Praxis versteht, in der Techniker_innen und Informatiker_innen ihre Vorstellungen von einem „guten Alter“ einschreiben und damit bestimmen, was altersgerechte Technik ist. Im Anschluss an Lindenberger verstehe ich AAL dabei erstens als Benutzerschnittstelle, die adaptiv und nutzerzentriert gestaltet ist und situativ und intelligent funktioniert. Zweitens ist AAL Mensch-Technik-Interaktion, die Selbständigkeit, Information und

allianzen/Ambient_Assisted_Living.html (letzter Zugriff: 30.06.2014)

[6] Diese Richtung wird auch als „Gerontechnology“ oder „Gerontotechnik“ bezeichnet und verknüpft Erkenntnisse, Paradigmen und Methoden der Human-Factors-Forschung mit Ansätzen und Entwicklungen der „Assistive Technology“ unter dem Paradigma altersgerechter Anwendung [Ch01; Bo07; Bo09; KT13].

Kommunikation, Wohnen, Gesundheit und Sicherheit in Handlung zwischen menschlichen und nicht-menschlichen Akteuren übersetzt und damit Handlungspotentiale nicht nur erzeugt, sondern vor allem auch verteilt bzw. zuweist. Diese Praktiken der Delegation müssen als machtvoll verstanden werden. Drittens ist AAL aber auch ein Innovationsfeld, in dem neue interaktive bzw. intelligente Systemtechnologien entwickelt werden sollen, um – viertens – den Anforderungen demographischer Entwicklungen begegnen zu können. Die argumentative Verschränkung von technischer zugunsten demographischer Entwicklung muss als eine politische Strategie verstanden werden, dem zukünftigen Pflegebedarf mittels Technik zu begegnen und gleichzeitig, wenn auch eher verdeckt, „Wirtschaftsförderung zu betreiben", so wie das von einem meiner Interviewpartner formuliert wird (Interview, April 2014). Dabei bleibt offen, inwieweit der Einsatz technischer Assistenzsysteme das soziale Handeln älterer Menschen und ihre kulturelle Praxis des Alterns verändern wird. Ich verstehe diese Frage auch als politische Frage, die im Diskurs um altersgerechte Technologien meist unsichtbar gemacht wird. So ist die erfolgreiche Nutzung an eine Vielzahl nicht nur technischer sondern vor allem sozialer Bedingungen gebunden, die in den technologieorientierten AAL-Projekten bisher nur wenig Berücksichtigung findet. Dieser blinde Fleck wiegt umso schwerer, als dass hier Technologien für eine Zielgruppe entwickelt werden, deren inter- und intraindividuelle Heterogenität sowohl hohe technische als auch ökonomische und nicht zuletzt ethische und rechtliche Ansprüche an die Entwicklung und Produktion technischer Lösungen stellt, die Aspekte sozialer Ungleichheit ebenso wie die Sicherstellung persönlicher Integrität und Autonomie adressiert [Ma12]. In den Projekten findet häufig eine Verkürzung dieser Problematik statt, die mit dem Verweis, dass es sich hier um ethische Problemlagen handelt, auf die im begrenzten Projektzeitraum und aufgrund der fehlenden Expertise der Projektbeteiligten nur eingeschränkt reagiert werden kann, reagiert wird. Dieser Begründung setze ich die Annahme entgegen, dass diese Aspekte im Sinne Latours gezielt geblackboxt werden [La02]. Die Beteiligungspraktiken von Nutzer_innen exemplifizieren diese Logik beispielhaft und sollen im Folgenden Gegenstand meiner Ausführungen sein.

2 *Users matter* – Wie die Nutzer_innen in die Technik kommen

Als erstes absolviert Herr Kastner das Musik Memory.[7] „Heute geht es aber wirklich schwerer als sonst", meint Herr Kastner. Auch bei der nächsten Aufgabe gibt es Schwierigkeiten, das Programm überspringt einzelne Aufgaben oder ändert während der Aufgabenbearbeitung die Bildschirmoberfläche, so dass die Aufgaben nicht abschließend gelöst werden. „Jetzt, sehen Sie, jetzt ist der so gesprungen!" Herr Kastner zeigt auf den Bildschirm. „Das macht der immer mal und man weiß gar nicht warum." Da erscheint schon die nächste Aufgabe und Herr Kastner tippt wieder das falsche Feld

[7] Folgende Ausführungen zu Herrn Kastner (anonymisiert) beruhen auf meiner Feldforschung in einem AAL-Projekt, welche ich seit März 2014 durchführe und im Dezember dieses Jahres abschließen werde.

an. „Ja, das bringt einen dann raus“, sagt er wütend, nachdem er drei Aufgaben hintereinander nicht richtig lösen konnte, „das ärgert einen dann schon oder man möchte wissen, warum das so ist“.

Folgt man den Beobachtungen von Nelly Oudshoorn und Trevor Pinch, so könnte man behaupten, dass es in den letzten Jahren eine Art *user turn* in den Technikwissenschaften gegeben hat, in dem die Maxime „users matter“ evident geworden ist [OP08]. Gerade die Diskussion um die Frage des „social shaping of technology“ [MW85] hat zu einer Fokussierung auf Aspekte der Nutzer_innenkonstruktion und -beteiligung geführt [OP03; ORS04]. Gegenstand dieser vornehmlich aus dem Kontext der *Science and Technology Studies* kommenden Studien, war die Frage, wie Nutzer_innen in Technologieprozessen konfiguriert werden [Wo91]. Dabei spielte bereits in der frühen Phase der Hinwendung zum „User“ die Frage der Repräsentation eine Rolle. „While ‘lead users’ often self-identify, there is clearly an issue about how users with no voice are represented. […] This raises the interesting issue that users may represent other groups as end-users while at the same time promoting their own interests.“ [OP08]

In Bezug auf den hier interessierenden Kontext der Anwendung technischer Assistenzsysteme durch ältere Nutzer_innen scheint die aus der feministischen Soziologie stammende Unterscheidung in „end-users“ [CC985], „lay end-users“ [SOK00] und „implicated actors“ [Cl98] produktiv. Vor allem das von Adele Clarke vorgeschlagene Konzept der „implicated actors“ verweist auf die ambivalente Repräsentation von Nutzer_innen in AAL-Projekten. Clarke versteht unter „implicated actors“, diejenigen welche von der Technik beeinflusst und diskursiv als „die Nutzer_innen“ konstruiert und angesprochen werden, im Entwicklungsprozess aber stumm oder physisch nicht anwesend sind [Cl98]. Während der von Clarke und anderen vertretene feministische Ansatz vor allem auf Fragen der Verteilung von Macht und der Anerkennung von Diversität beruht, wurden in den *Science and Technology Studies* vor allem semiotische Ansätze herausgearbeitet, die stärker aus einer wissenssoziologischen Perspektive die Frage stellen, „how meanings are built“ [OP08]. Steve Woolgar hat dabei die Konfigurierung der Nutzer_innen im Designprozess analysiert, um zu betonen, dass die Art und Weise wie Nutzer_innen technische Geräte benutzen, bereits durch Design- und Produktionsprozesse festgelegt ist [Wo91]. Diese Konfigurierungsarbeit wurde in der Folge auf weitere Akteure erweitert, die an dieser Arbeit beteiligt sind [Ma00]. „From this perspective, technological development emerges as a culturally contested zone where users, patient advocacy groups, consumer organizations, designers, producers, salespeople, policymakers, and intermediary groups create, negotiate, and give differing, sometimes conflicting forms, meanings, and uses to technologies“ [OP03]. Auch Akrich und Latour gehen mit ihrem Konzept des “scripts” davon aus, dass sich in der Herstellung und Gestaltung neuer (technischer) Produkte die Vorstellungen der am Produktionsprozess Beteiligten, wie Ingenieur_innen, Designer_innen oder Techniker_innen, über zukünftige Nutzer_innen in den Objekten materialisieren [AL92; Ak92; Ak95]. Diese Objekte stellen in Folge neue oder transformierte „geographies of responsibilities“ her [Ak92]. Im Gegensatz zu Woolgar verstehen Akrich und Latour den Prozess des Konfigurierens/Einschreibens nicht als unidirektionalen linearen Prozess,

indem die Nutzer_innen passiv bleiben, vielmehr schlägt Akrich mit „subscription“ und „de-inscription“ [Ak92] Begriffe vor, die für eine praxeologisch-ethnografische Beschreibung der Nutzer_innenpraktiken hilfreich sind.

Auch bei der nächsten Übung, dem Zahlenvergleich, überspringt das Programm Aufgaben. „Eigentlich stört mich das jetzt nicht so“, kommentiert Herr Kastner das Vorgehen, „aber ich frage mich, wie das dann mit den Punkten ist, wie das dann gezählt wird, weil ich die Aufgabe ja gar nicht machen konnte. Da fehlen mir dann ja Punkte“. Herr Kastner trainiert seit vier Wochen an einem Tablet-PC sein Gedächtnis mit Hilfe einer Plattform, die ihm dazu aus einem konstanten Set aus zehn kognitiven Aufgabenformaten fünf Aufgabenformate auswählt, die Herr Kastner nach einer vorgegeben Reihenfolge und in einer bestimmten Zeit absolvieren muss. Die Punkte, die Herr Kastner bei jeder Übung erreicht, werden aufsummiert und ihm als persönlicher Score mittels einer graphischen Darstellung veranschaulicht. Die graphische Darstellung besteht aus einem Kurvendiagramm und einer Tabelle, in welcher grüne Pfeile eine Verbesserung und rote Pfeile eine Verschlechterung der Trainingsergebnisse symbolisieren. Das Kurvendiagramm erscheint nach jeder Übung, die Pfeile wenn Herr Kastner alle fünf Übungen seines Trainingsplans vollständig absolviert hat. Kommt es zu technischen Problemen, beispielsweise zu Unterbrechungen der WLAN-Verbindung oder zu einer Überlastung des Programms, bricht dieses unerwartet ab und Herr Kastner muss mit einem neuen Trainingsplan beginnen. Um einem Verlust von Punkten vorzubeugen, zählt Herr Kastner aus diesem Grund, die nach jeder Übung angezeigten Punkte selbst mit, wobei er festgestellt hat, dass die Punkte, welche er zählt, nicht mit den Punkten, die das System verrechnet, übereinstimmen. „Ich weiß das ja jetzt, dass das nicht stimmt“, erklärt mir Herr Kastner seine Kontrollmethode, „aber ich frage mich, wie das dann anderen geht, denen das nicht auffällt und die ja dann auch so ein Endergebnis bekommen“. Er teilt mir diese Beobachtung mit, ich schreibe diese Beobachtung auf, zusätzlich hält er sie auf dem Fragebogen, den ihm das Projekt zur Evaluation zugeschickt hat fest. Ich werde seine Beobachtung an das Projektteam rückmelden und Herr Kastners Fragebogen wird ausgewertet. Zu welcher Konsequenz seine Beobachtung aber führen wird, das wissen weder er noch ich.

Nutzer_innenbeteiligung in AAL-Projekten folgt dem Ansatz des *User-Centered Design* (UCD), das im deutschen Kontext auch unter dem Begriff *menschzentrierte Gestaltung* firmiert. Dieses ISO-Norm geprüfte Verfahren zielt darauf ab, Systeme gebrauchstauglich und zweckdienlich zu gestalten, indem der Bedarf und die Anforderungen potentieller Benutzer_innen berücksichtigt werden. Dadurch sollen die Effektivität, Effizienz und Zufriedenstellung erhöht und die Zugänglichkeit und Nachhaltigkeit der Technologien verbessert werden. Für staatlich geförderte AAL-Projekte ist dieser Ansatz verpflichtend. Eine Verpflichtung, die der praktischen Realisierung viel Handlungsspielraum lässt. So bezeichnete ein renommierter Usability-Consultant, der selbst Technologieprojekte durchführt, begleitet und berät, die Umsetzung dieses Anspruchs in den Projekten als „eine Farce“ (Feldnotiz, März 2014). Auch die Mitarbeiterin in einem von mir beforschten Projekt betitelte die Verpflichtung auf Nutzer_innenbeteiligung als „Antragsprosa“ (Interview, April 2014). Diese

Aussagen weisen daraufhin, dass UCD etwas ist, das gemacht werden muss, dass erst in den Projekten durch die Beteiligung aller relevanten Akteur_innen möglich wird und durch deren Praktiken in die Technik übersetzt wird.

Dass dieser Anspruch im Alltag der Projekte zur „Farce“ wird, wirft die Frage auf, inwieweit der Anspruch nutzer_innenzentriert zu arbeiten, an die falschen Akteur_innen adressiert ist, nämlich Informatiker_innen und Ingenieur_innen, die weder dafür ausgebildet sind noch in ihrem professionellen Selbstverständnis darauf fokussieren, mit Nutzer_innen zu interagieren. Die Folge sind Alibi-Praktiken und Übersetzungsprobleme. Hier stellt sich die Frage, inwieweit die Förderinstitutionen diese Alibi-Praktiken unterstützen und damit ein System aufrecht erhalten, dass an den tatsächlichen Bedürfnissen der Nutzer_innen vorbei entwickelt und trotz zahlreicher Beispiele verkennt, dass Nutzer_innenbeteiligung gezielter Koordination, Steuerung und Training bedarf.

An dieser Stelle wird auch deutlich, dass die Entwicklung „[a]ltersgerechte[r] Assistenzsysteme für ein gesundes und unabhängiges Leben“ (BMBF) sich nicht auf die Entwicklung von „Konzepte[n], Dienstleistungen und Produkte[n], die neue Technologien und soziales Umfeld miteinander verbinden, um die Lebensqualität für Menschen in allen Lebensabschnitten zu erhöhen“ (VDE/VDI-IT) begrenzen lässt, sondern als eine politische Praxis verstanden werden muss, mit der staatlich gefördert Interventionen in sensible Lebensbereiche wie Gesundheit und Pflege ermöglicht und legitimiert werden. Technisierungen des Alltagshandelns, die einhergehen mit zahlreichen sozialen und kulturellen Veränderungen, entstehen eben erst durch diese Technologien. „This means […]“, so Akrich, „that new technologies may not only lead to new arrangements of people and things. They may, in addition, generate and ‘naturalize’ new forms and orders of causality and, indeed, new forms of knowledge about the world.“ [Ak92].

3 *Doing Users* - Nutzer_innenbeteiligung als *tipping point* in AAL

Man muss schon den richtigen Einstieg finden“, meint Herr Schreiner, als er von seinen Projekterfahrungen berichtet, „entweder [trifft man] auf Personen mit einer sowieso hohen Motivation, allein aus Neugier, während bei der anderen Gruppe Interesse über persönliche Vorlieben geweckt wird“ (Feldnotiz, Mai 2013). Herr Schreiner sitzt neben Herrn Wolf und verfolgt, wie dieser die Plattform testet, die Herr Schreiner mit seinen Projektpartner_innen seit zwei Jahren entwickelt. „Zuhause würde ich jetzt einfach anrufen“, sagt Herr Wolf, „ich würde mein Telefon nehmen und dann einfach den anrufen, mit dem ich das machen will, und dann weiß ich auch gleich, ob der kann oder nicht.“ Aber ein Telefon liegt nicht auf dem Tisch des Usability-Labors, also versucht Herr Wolf die Anwendung zu öffnen. „Knopf drücken und dann rüberschieben“, murmelt er, während er den weißen Knopf drückt und den auf dem Bildschirm sichtbar werdenden Pfeil nach rechts schiebt. Aber nichts passiert, der Pfeil schnippt zurück in

die Ausgangsposition. Herr Wolf wackelt auf seinem Stuhl, rückt ein Stück näher und probiert es erneut. Es funktioniert, der Pfeil verschwindet und ein neuer Bildschirm öffnet sich. „Das ist eben das Tolle am CaseTab, man nimmt es in die Hand, schaltet es ein und los geht`s!“, kommentiert Herr Schreiner die Szene.

Wie bedarfsgerecht und nutzer_innenfreundlich die Gestaltung des technischen Produkts sich vollzieht, so zeigt diese Szene, ist von der Art und Weise wie Beteiligung gemacht wird, welche Personen dazu ausgewählt werden und mit welcher Intention die Beteiligung stattfindet, abhängig. Hier spielen Fragen der Auswahl und Beteiligung ebenso eine Rolle, wie die Frage nach den Ressourcen, über die das jeweilige Projekt verfügt, Nutzer_innen im Sinne des UCD zu beteiligen.

Im Falle BMBF-geförderter AAL-Projekte ist die häufigste Projektform ein Verbund unterschiedlicher Träger aus den Sektoren Technik, Soziales und Wissenschaft. Je nach Projektgröße variiert die Zahl der Partner durchschnittlich zwischen fünf und acht.[8] Bereits in der Beschreibung der Arbeitspakete im Rahmen der Antragstellung werden die Aufgaben den jeweiligen Partnern zugeordnet. Diese Trennung bleibt auch im Projektverlauf bestehen, wie mir eine AAL-Beraterin berichtete: „Das hat man eigentlich von Anfang an fest und dann macht jeder seins und die sehen sich dann vielleicht mal alle halbe Jahre zum Zwischenbericht oder so“ (Feldnotiz, Juni 2014).

Die Durchführung des UCD fällt meist einem der beteiligten Partnern zu und stellt somit kein Instrument dar, mit welchem innerhalb der Projekte andere Formen der Zusammenarbeit etabliert werden könnten. Der entsprechende Partner hat dann die Aufgabe über Anforderungs- und Bedarfsanalysen zu Beginn des Projekts, über formative Evaluationen innerhalb des iterativen Entwicklungsprozesses bis hin zu summativen Evaluationen, den sogenannten Nutzertests, die Perspektive potentieller Nutzer_innen einzubringen. Häufig verfügt der durchführende Projektpartner nicht über spezifische Kenntnisse der Usability Studies und eignet sich die einzelnen Aufgabenbereiche im Rahmen des Projekts an. In selteneren Fällen wird das Arbeitspaket UCD tatsächlich von professionellen Usability-Consultants oder Wissenschaftler_innen aus dem Bereich Human Factors/Kognitive Ergonomie durchgeführt. Das heißt, wie Nutzer_innenbeteiligung realisiert wird, beruht auf Entscheidungen nichtprofessioneller Usability-Praktiker_innen, auf deren Vorerfahrungen und methodischen Kenntnissen sowie auf deren Bereitschaft sich mit UCD-Methoden und Zugängen vertraut zu machen. UCD erscheint somit vor allem als ein praxisgeleitetes Handeln, das stark von den personellen Ressourcen desjeweiligen Projektpartners abhängt.

Im Beispiel eines AAL-Projekts aus der 1. Bekanntmachung des BMBF „Selbstbestimmt Leben im Alter“[9] liegt die Nutzer_innenbeteiligung in den Händen eines Projektpartners (Verband von Wohnungsbaugenossenschaften), in dessen Kontext die Geräte später auch

[8] In europäischen Projektverbünden können es auch deutlich mehr Partner werden, dann liegt die Zahl zwischen zehn bis vierzehn im Durchschnitt.

[9] Siehe http://www.mtidw.de/ueberblick-bekanntmachungen/selbstbestimmt-leben (letzter Zugriff: 30.06.2014)

angewendet werden sollen. Die Nutzer_innen werden ausschließlich aus dem Bestand der Wohnungsbaugenossenschaft akquiriert und im Namen dieser auch die Tests durchgeführt. Trotz der fehlenden professionellen Expertise ist die Nutzer_inneneinbindung relativ umfangreich und stößt aufgrund der Vertrautheit zwischen Wohnungsbaugenossenschaft und Genossenschaftler_innen auf breite Resonanz. Neben einer Nutzer_innenbefragung mittels Fragebogen zu Bedarf und Anforderungen an unterstützende Technik im Wohnkontext, wurde ein Sozialfragebogen verteilt, um sowohl die finanzielle Bereitschaft der Genossenschafler_innen abschätzen zu können als auch ihre sozio-ökonomischen Lebensumstände besser zu verstehen und zu berücksichtigen. Darüber hinaus wurden Studien zu Sozialrendite, Farbgestaltung und Gesundheit in Auftrag gegeben und die Mitglieder in der vierteljährlich erscheinenden Mitgliederinformation regelmäßig über den Projektverlauf informiert. Zusätzlich gab es Veranstaltungen und Vorträge, ein Kontaktbüro vor Ort und eine Internetpräsenz sowie die Möglichkeit in der technisch altersgerecht gestalteten Wohnung der Wohnungsbaugenossenschaft auf Probe zu wohnen. All diese Strategien stießen bei den Befragten auf großes Interesse, so der Projektleiter in einem Interview (Januar 2014). Die Einführung der Technik in ein stabiles Wohnsetting und die kontinuierliche Weitergabe von Informationen über den Projektverlauf an die Genossenschaftler_innen beförderte die Beteiligung und die Bereitschaft sich auf die Technik einzulassen. Gerade die Anwesenheit und das persönliche Engagement der Projektvertreter_innen, das Kontaktbüro vor Ort und die vielen dezentralen Veranstaltungen wirkten sich förderlich auf die Akzeptanz und die erfolgreiche Realisierung des Projekts aus.

Die persönliche Anwesenheit der Projektpartner und ihr Engagement im Feld garantierten auch in einem Projekt aus der 2. Bekanntmachung „Mobil bis ins hohe Alter“[10] die Bereitschaft zur aktiven Beteiligung der Nutzer_innen, die wiederrum die spätere Akzeptanz und erfolgreiche Integration der Technik in die Lebenswelt der Nutzer_innen stabilisierten. In diesem zweiten Projekt, in welchem gemäß der Bekanntmachung nicht nur die Nutzer_inneneinbindung verpflichtend war, sondern auch die geistes- oder sozialwissenschaftliche Begleitforschung, wurden die Nutzer_innen regelmäßig von einem konstanten Kreis von Akteur_innen aus Wissenschaft, Stadtverwaltung und Pflegedienstleistung begleitet. Neben niedrigschwelligen Informationsveranstaltungen in Senior_innentreffs zu Beginn des Projekts beförderten vor allem Beteiligungsformate im Projektverlauf wie regelmäßig stattfindende Nutzer-Cafés oder Schulungen durch Senior_innen und/oder Schüler_innen sowohl die Nutzung der Plattform als auch eine aktive Beteiligung. So gab es beispielsweise selbstorganisierte informelle Treffen der Nutzer_innengruppe zum gemeinsamen Testen, Lernen und Austauschen sowie eine durch das Gerät vermittelte Kommunikation in Form einer Messenger-Gruppe. Parallel wurden die Teilnehmer_innen regelmäßig durch Sozialwissenschaftler_innen zu ihrem Nutzungsverhalten befragt, Nutzertests durchgeführt und Prototypen evaluiert, wobei das begleitende Team sich stets aus den gleichen Personen zusammensetzte. Die zeitnahen, regelmäßigen Treffen, der enge

[10] Siehe http://www.mtidw.de/ueberblick-bekanntmachungen/mobil-bis-ins-hohe-alter (letzter Zugriff: 30.06.2014)

Kontakt und die direkte persönliche Ansprache erhöhten bei den Nutzer_innen nicht nur die Bereitschaft die Plattform regelmäßig zu nutzen bzw. zu testen sondern auch die Bereitschaft sich kritisch mit Fragen von Mobilität und räumlicher Infrastruktur in ländlichen Gebieten auseinanderzusetzen. Während die meisten zu Beginn der Studie PKW-Nutzer_innen waren, so führte die Teilnahme nicht nur theoretisch sondern auch praktisch zur Nutzung des regionalen ÖPNV, wodurch die Nutzer_innen ein Problembewusstsein für die Belange anderer Älterer entwickelten, die zum Beispiel über kein eigenes PKW verfügen. Die Anwesenheit von Vertreter_innen der Projektpartner wurde dabei nicht als störend oder kontrollierend sondern vielmehr als Wertschätzung ihres Engagements als Nutzer_innen erlebt. Auch nach zwei Jahren waren alle 23 Teilnehmer_innen noch dabei.

Ähnlich zu den beiden eben genannten Projekten verlief auch die Nutzer_innenbeteiligung in einem dritten Projekt, wobei hier nicht die Wohnungsbaugenossenschaft oder der räumliche Kontext den gemeinsamen Rahmen bildete sondern die Anwendung geriatrisch-gerontologischen Wissens am Beispiel von kognitivem Training. Hier war es vor allem die Integration von wissenschaftlichem Wissen aus einem für das Alter(n) spezifischen Bereich, die eine Herausforderung für die Projektpartner darstellte. Während in den beiden erstgenannten Gruppen vor allem Laien wie Bewohner_innen oder Einheimische die Nutzer_innengruppe bildeten, wurde der Kreis der Nutzer_innen im Entwicklungsprozess je nach iterativer Stufe auf weitere Akteursgruppen wie Mediziner_innen, Pflegedienstleister_innen, Informatiker_innen und Sozialwissenschaftler_innen ausgeweitet. Zwar bildete die Evaluation durch die Nutzer_innen auch hier den zentralen Bestandteil des UCD, doch wurde dieser von weiteren Formen der Evaluation ergänzt, die vor allem auf die wissenschaftliche Fundierung der Projektergebnisse zielte. Dabei war es gerade diese wissenschaftliche Kontextualisierung der Trainingselemente und des Informationsmaterials, die bei den Nutzer_innen zu Vertrauen und aktiver Teilnahme führten. Auch wenn es dabei zu keinem regelmäßigen Austausch zwischen „den“ Nutzer_innen und „den“ Tester_innen kam, wie im zweiten Projekt, so beförderte die wissenschaftliche Rahmung des Programms die Seriosität desselben und motivierte die Teilnehmer_innen zu ernsthaftem engagiertem Training.

Diese drei Projekte geben einen Eindruck, wie ältere Menschen zu Nutzer_innen von Technologien gemacht werden, die sie überhaupt erst als solche, nämlich „ältere Nutzer_innen assistiver Technologien“ adressieren. Sie zeigen auch, dass die Vorstellung, die Akquise von Proband_innen mit einem Alter 60+, deren Kontaktdaten in einer Datenbank liegen und die es nur gilt anzuschreiben, einzuladen und zu testen, dem gedanklichen Trugschluss obliegt, dass die Nutzer_innen in den Projekten einen natürlichen Ausschnitt gesellschaftlicher Wirklichkeit repräsentieren. Nutzer_innen in AAL sind etwas Gemachtes, die erst in der Durchführung des UCD, sozusagen *in the making*, entstehen. Erst die Ansprache als Nutzer_innen, ihre Beteiligung am Entwicklungsprozess und die Delegation von Handlungspotential im Sinne Latours (1992) lässt die Adressierten zu Nutzer_innen werden. Sie stellen damit nicht nur eine künstliche Auswahl dar, die eine natürliche Entität suggerieren soll, sondern sie sind

eben vor allem eine Projektion – ein Bild von einem älteren Nutzer.

Die Frage nach der Beteiligung von Alltagsexpert_innen an Entwicklungsprozessen und die Frage nach dem Zusammenspiel dieser Alltagsexpert_innen mit Fach-Expert_innen lässt sich dabei parallel zu der Frage lesen, welche die beiden Wissenschaftssoziologen Harry Collins und Robert Evans in Bezug auf die Beziehung von Wissenschaft und Politik aufwerfen. Sie gehen davon aus, dass nicht nur Wissenschaftler_innen Wissen produzieren, das auf Gesellschaft bzw. Politik zurückwirkt, sondern auch sogenannte Laienexpert_innen. Deshalb stellt sich ihnen die Frage, wie weit der Kreis der teilnehmenden Akteur_innen ausgeweitet werden kann und muss. Sie bezeichnen dieses Problem als „problem of extension" [CE02]. Estrid Sørensen antwortet auf das von Collins und Evans aufgeworfene Problem der Ausweitung mit der Aussage, dass die „Ausweitung auf *relevante* Betroffene […] politisch entschieden werden" muss [Sø12]. Die Entwicklung altersgerechter Assistenzsysteme kann folglich als eine politische Praxis verstanden werden, in der die Beteiligung von Alltagsexpert_innen zum *tipping point* für die erfolgreiche Entwicklung der Produkte, ihre Marktreife und ihre Integration in die Lebenswelt älterer Nutzer_innen wird.

4 Nutzer_innenbeteiligung als Ko-Konstruktion von Alter(n) und Technik

An dieser Stelle erscheint die Erweiterung einer techniksoziologischen Lesart um Ansätze aus den Gender Studies vielversprechend, um innerhalb der Forschungsfrage, wie das „Alter(n)" in AAL-Technologien eingeschrieben wird zu konturieren. Dabei kann sowohl auf die Arbeiten Judy Wajcmans zur Ko-Konstruktion von Technik und Geschlecht in Technikforschungsprozessen rekurriert werden [Wa02] als auch auf das Konzept des „Genderskripts" wie es Nelly Oudshoorn, Ellen van Oost und Els Rommes zur Beschreibung von strukturellen, symbolischen und identitären Praktiken der Vergeschlechtlichung von technischen Artefakten vorschlagen [ROO99]. Der Genderskript-Ansatz beruht dabei auf dem Begriff des „Skripts" wie ihn Madelaine Akrich entwickelt hat, um die Designprozesse von Entwickler_innen als Einschreibungsprozesse sichtbar zu machen [Ak92]. Ausgangspunkt von Akrichs Überlegungen ist dabei die Beobachtung, dass Entwickler_innen Vorstellungen über den Nutzungskontext, die Nutzungspraktiken und vor allem über die Nutzer_innen während des Entwicklungsprozesses produzieren und diese die Designentscheidungen und den Gestaltungsprozess beeinflussen: „A large part of the work of innovators is that of ‚*inscribing*' this vision of (or prediction about the world in the technical content of the new object." [Ak92].

Ausgehend von dem Genderskript-Ansatz möchte ich den Begriff des „Ageskripts" einführen, um die Ko-Konstruktion von Alter und Technik zu beschreiben. Indem Entwickler_innen altersgerechte Technik herstellen, schreiben sie nicht nur ihre Vorstellungen von Alter(n) und altersgerechter Techniknutzung in die Technik ein,

sondern konstruieren gleichzeitig auch ein technikgerechtes Alter(n), welches ebenfalls in die Technologien eingeschrieben wird. Dieser doppelte Einschreibungsprozess kann mittels des „Ageskript"-Ansatzes sichtbar gemacht werden.

AAL kann dann als eine soziale Technologie verstanden werden, die erst durch das Aushandeln unterschiedlicher Akteur_innen entsteht. Zu diesen interagierenden Akteur_innen zählen neben den technischen Geräten und ihrer benötigten Infrastruktur vor allem die Entwickler_innen und Nutzer_innen, die folgt man dem Ideal des UCD in einem engen iterativen Prozess miteinander aushandeln, wie die Technik gestaltet werden soll. Neben diesen sind aber auch weitere Stakeholder am Aus-Agieren technischer Entwicklung beteiligt und müssen als ko-konstruierend verstanden werden, wie zum Beispiel Angehörige, (Pflege)-Dienstleister_innen, politische Förderinstitutionen und nicht zuletzt die beobachtende Ethnographin. Sie alle handeln aus, wie in dem konkreten Projekt, mit der spezifischen Technik – Tablet oder Armband, Teppich oder Brille – Autonomie und Teilhabe hergestellt werden. Ko-Konstruktion von Technik und Alter in AAL-Projekten ist dabei oftmals transitiv und prozesshaft und nicht zu gleichen Teilen sinnstiftend für das Handeln der Akteur_innen. So zeigen beispielsweise indem von mir untersuchten Projekt zum kognitiven Training die technischen Entwickler_innen wenig Interesse an den Beiträgen der Nutzer_innen. Diese Anerkennungsproblematik tritt auch innerhalb des Projekts auf. So beschreibt eine Projektmitarbeiterin den mühsamen Prozess, indem sie versucht ihre kognitionswissenschaftlichen Vorschläge zur Gestaltung eines Gedächtnistrainings den ausführenden Techniker_innen verständlich, transparent und anwendbar zu machen: „Da hab ich zum Beispiel dann auch angefangen so Zeichnungen zu machen, die liegen alle in dem Ordner, damit die dann sehen, wie ich mir das vorstelle, aber dann habe ich gemerkt, dass da nie jemand reinguckt. [lacht] Das hab ich dann wohl eher so für mich gemacht?" (Interview, April 2014). Die Übersetzungsarbeit, welche die Projektmitarbeiterin leistet, stellt den Versuch dar, an der Einschreibung von Nutzer_innenvorstellungen teilzuhaben und diese mitzubestimmen, da sie selbst aber nur „Lieferantin" von Wissen ist und als solche von ihren technischen Projektpartner_innen angesprochen wird, entsteht wie Moser und Law sagen würden ein Artikulationsproblem [Ak92]. Die Projektmitarbeiterin erlebt ihre hegemoniale Marginalisierung innerhalb des Projekts kongruent zu den Erfahrungen der Nutzer_innen, deren Punktwerte nicht mehr stimmen.

Das heißt, Nutzer_innenbeteiligung ist relational, nicht nur die teilnehmenden älteren Proband_innen sind wie Moser und Law argumentieren „constructed in relations" [ML03], sondern auch die Mitarbeiter_innen in den Projekten und nicht zuletzt die Technik selbst. Im Verlauf der Zusammenarbeit bilden diese Akteure ein Arrangement aus Relationen, das aus der Perspektive der Actor-Network-Theory als ein soziotechnisches Netz oder Ensemble verstanden werden kann, in dem die Akteure ein spezifisches Wissen, eine spezifische Praxis und eine spezifische Macht herstellen, einschreiben, teilen und dabei AAL, verstanden als eine materialisierte Mensch-Technik-Interaktion, erfinden, gestalten und stabilisieren. Die Praxis der Nutzer_innenbeteiligung muss dann verstanden werden als eine Praxis, die eingebunden ist in eine Pragmatik der

Relationen und Strukturen, indem die jeweiligen Akteur_innen je nach Situation über ganz unterschiedliche Handlungspotentiale verfügen und diese aktiv einsetzen oder auch verweigern. Während der Usability-Consultant über vermeintlich viel Gestaltungsmacht qua seines Status als Projektleiter verfügt, wird in der Beobachtung deutlich, dass auch sein Handeln eingebunden ist in das Handeln von Förderprogrammen und Finanzierungsmodellen. Das heißt auch, dass es in der Beobachtung von AAL aus einer kritischen möglicherweise sogar politischen Perspektive redundant wäre Experten- und Laienstatus als fix gesetzte Größen in den verschiedenen Akteurssettings zu verstehen. Vielmehr ist es die Situation und der dadurch aufgerufene Kontext, der Handlungsmacht verteilt, Rollen zuweist und Entscheidungen fallen lässt. Welcher Inhalt letztendlich in das Gerät eingeschrieben wird bzw. wie das Gerät konstruiert wird, bestimmt ein heterogenes Geflecht an sichtbaren und unsichtbaren Anforderungen und Widerständen, menschlichen und nicht-menschlichen Akteur_innen, die sowohl durch die technischen Möglichkeiten, die finanzielle Ausstattung aber vor allem durch das gegenseitige Verständnis und eine gemeinsame Auffassung dessen, was altersgerechte Technik in dem jeweiligen Nutzungskontext – Wohnen, Mobilität oder Kognition – sein soll und sein kann, bestimmt wird.[11] Sowohl die Techniker_innen, als auch die begleitenden Wissenschaftler_innen, beratenden Dienstleister_innen, testenden Nutzer_innen und fördernden Institutionen müssen sich als handelnde Akteure im Gestaltungsprozess wahrnehmen und diese Akteur_innenposition auch als eine der Verantwortlichkeit begreifen, die sich eben nur begrenzt delegieren lässt. Dass ein solches Verantwortlichkeitsensemble Steuerung, Beratung und Kontrolle bedarf, fand bisher in keiner Bekanntmachung Berücksichtigung.

Der Entwicklungsprozess von AAL-Technologien gleicht damit einem Arrangement, das die beteiligten Akteur_innen miteinander aushandeln. Dazu zählen Informatiker_innen, Designer_innen, Wissenschaftler_innen, Projektträger_innen, Nutzer_innen, Geräte und Infrastrukturen. In diesem Arrangement wird User-Centered Design dann zu einer kooperativen Praxis – transparent, partizipativ, gleichberechtigt – wenn ältere Nutzer_innen frühzeitig und ernsthaft in die Entwicklung einbezogen werden. Um ein solches Arrangement zu erreichen, erscheint ein Paradigmenwechsel lohnenswert, der die Nutzerbeteiligung öffnet und alle Projektbeteiligten zu Nutzer_innen macht.

Literaturverzeichnis

[Ak92] Akrich, M.: The De-Scription of Technical Objects. In W. E. Bijker, & J. Law (Hrsg.) Shaping Technology / Building Society: Studies in Sociotechnical Change (S. 205-225).

[11] Die Marginalisierung von Mitsprachepositionen durch Praktiken des Abwartens, Verzögerns, Verweigerns oder Kontrollierens innerhalb des Projektnetzwerks kann auf Seiten der Marginalisierten zu widerständigen oder resignativen Praktiken führen. Die Förderstruktur lässt jedoch kein Aushandeln dieser Spannungen zu, es gibt weder die Möglichkeit der Beschwerde noch des Aufkündigens der Zusammenarbeit ohne den Verlust der eigenen Projektbeteiligung. Auf diese Weise werden Kompromisse produziert, die die Gestaltung der Technik zum einen sowohl an der zeitlichen als auch technischen Machbarkeit des Projekts ausrichten, zum anderen aber auch an den Anforderungen an und der Bereitschaft zu einer Projektzusammenarbeit.

Cambridge, MA: MIT Press. 1992.

[AL 92] Akrich, M. & Latour, B.: A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies. In W. Bijker, & J. Law (Hrsg.) Shaping Technology—Building Society: Studies in Sociotechnical Change (S. 259–64). Cambridge, MA: MIT Press, 1992.

[Ak95] Akrich, M.: User Representations: Practices, Methods and Sociology. In A. Rip, Th. J. Misa, & J. W. Schot (Hrsg.) Managing Technology in society: the approach of constructive technology assessment (S. 167-184). London: Pinter, 1995.

[Bo07] Bouma, H. et al.: Gerontechnology in perspective. Gerontechnology, 6/4, 190-216, 2007.

[Bo09] Bouma, H. et al. : Gerontechnology as a field of endeavour. Gerontechnology 8/2, 68-75, 2009.

[CC98] Casper, M., & Clarke, A.: Making the Pap Smear into the Right Tool for the Job: Cervical Cancer Screening in the United States, c. 1940–1995. Social Studies of Science 28/2, 255–90, 1998.

[Ch01] Charness, N. et al.: Communication, Technology and Aging: Opportunities and challenges for the future. New York, NY: Springer, 2001.

[Cl98] Clarke, A.: Disciplining Reproduction: Modernity, American Life and 'The Problem of Sex'. Chicago: University of Chicago Press, 1998.

[Cl05] Clark, A.:. Situational Analysis – Grounded Theory after the Postmodern Turn. Thousand Oaks: Sage Publications, 2005.

[CE02] Collins, H., & Evans, R.: The Third Wave of Science Studies. Studies of Expertise and Experience. The Social Studies of Science, 32/2, 235-296, 2002.

[Cz97] Czaja, S. J.: Computer technology and the older adult. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Hrsg.) Handbook of Human–Computer Interaction (S. 797-812). Amsterdam: Elsevier, 1997.

[CS98] Czaja, S. J., & Sharit, J.: Age differences in attitudes towards computers: The influence of task characteristics. The Journals of Gerontology: Psychological Sciences and Social Sciences, 53/5, 329-340, 1998.

[CL07] Czaja, S. J., & Lee, C. C.: The impact of Aging on Access to Technology. Universal Access in the Information Society, 5/4, 341-349, 2007.

[DL12] Denninger, T., & Lessenich, St.: Rechtfertigungsordnungen des Alter(n)s, Soziale Welt, 63, 299-315. 2012.

[De14] Denninger, T., Dyk, S. v., Lessenich, St., & Richter, A.: Leben im Ruhestand: Zur Neuverhandlung des Alters in der Aktivgesellschaft. Bielefeld: transcript, 2014.

[DAP07] Dickinson, A., Arnott, J., & Prior, S.: Methods for human – computer interaction research with older people. Behaviour & Information Technology, 26/4, 343–352, 2007.

[DL09] Dyk, S. v., & Lessenich, St.: Die jungen Alten. Analysen einer neuen Sozialfigur. Frankfurt a.M.: Campus Verlag. 2009.

[FR97] Fisk, A. D, & Rogers, W. A.: Handbook of Human Factors and the Older Adult. San

Diego CA: Academic Press, 1997.

[Fi09] Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J.: Designing for Older Adults: Principles and Creative Human Factors Approaches. Boca Raton FL: CRC Press, 2009².

[Ha00] Hawthorn, D.: Possible implications of aging for interface designers. Interacting with computers, 12/5), 507-528, 2000.

[KC95] Kelly, C. L. & Charness, N.: Issues in training older adults to use computers. Behaviour and Information Technology, 14/2, 107-120, 1995.

[KMM99] Kubeck, J. E., Miller-Albrecht, S. A., & Murphy, M. D.. Finding information on the World Wide Web: Exploring older adult's exploration. Educational Gerontechnology, 25/2, 167–183, 1999.

[KT13] Künemund, H., & Tanschuss, N.: Gero-Technology: Old Age in The Electronic Jungle. In K. Komp, & M.Aartsen (Hrsg.) Old Age in Europe: A Textbook of Gerontology (S. 97-112). New York: Springer, 2013.

[La] Latour, B.: Die Hoffnung der Pandora. Untersuchungen zur Wirklichkeit der Wissenschaft. Frankfurt a.M.: Suhrkamp, 2002.

[Li11] Lindenberger, U., Nehmer, J., Steinhagen-Thiessen, E., Delius, J., & Schellenbach, M.: Altern und Technik: Altern in Deutschland, Band 6. Stuttgart: Wissenschaftliche Verlagsgesellschaft, 2011.

[Ma00] Mackay, H., C. Crane, P. Beynon-Davies, & D. Tudhope : Reconfiguring the User: Using Rapid Application Development. The Social Studies of Science 30, 737–59, 2000.

[MW85] MacKenzie, D., & Wajcman, J.: The Social Shaping of Technology. Milton Keynes, U.K.: Open University Press, 1985.

[Ma12] Manzeschke, A., Weber, K., Rother, E., & Fangerau, H.: Ethische Fragen im Bereich Altersgerechter Assistenzsysteme. Ergebnisse der Studie. Berlin, 2012.

[Me11] Medhi, I. et al. : Designing mobile interfaces for novice and low-literacy users. ACM Transactions on Computer-Human Interaction, 18/1, 1-28, 2011.

[Mi10] Mitzner, T. L. et al.: Older adults talk technology: Technology usage and attitudes. Computers in Human Behavior, 26/6, 1710-1721, 2010.

[ML03] Moser, I., & J. Law: 'Making Voices': New Media Technologies, Disabilities, and Articulation. In G. Liestol, A. Morrison, & T. Rasmussen (Hrsg.) Digital Media Revisited: Theoretical and Conceptual Innovation in Digital Domains (S. 491–520). Cambridge, MA: MIT Press, 2003.

[OP03] Oudshoorn, N., & Pinch, T.: Introduction. How Users and Non-users Matter. In N. Oudshoorn & T. Pinch (Hrsg.) How Users Matter: The Co-Construction of Users and Technologies (S. 1-25). Cambridge, MA: MIT Press, 2003.

[ORS04] Oudshoorn, N., Rommes, E., & Stienstra, M.: Configuring the User as Everybody: Gender and Cultures of Design in Information and Communication Technologies. Science, Technology & Human Values 29, 30–64, 2004.

[OP08] Oudshoorn, N., & Pinch, T.: User-Technology Relationships: Some Recent

Developments, 2008, Online Ressource: http://csid.unt.edu/research/trw/User-Technology%20Relationships,%20Some%20Recent%20Developments.pdf, [30.07.2014].

[ROO99] Rommes, E., Van Oost, E., & Oudshoorn, N.: Gender in the Design of the Digital city of Amsterdam. Information, Communication & Society 2/4, 476-495, 1999.

[Ro01] Rogers, W. A. et al. :A systems approach for training older adults to use technology. In N. Charness et al. (Hrsg.) Communication, Technology and Aging: Opportunities and challenges for the future (S 187-208). New York: Springer.

[SOK00] Saetnan, A., Oudshoorn, N., & Kirejczyk, M.: Bodies of Technology: Women's Involvement with Reproductive Medicine. Columbus: Ohio State University Press, 2000.

[Sc00] Schroeter, K. R.: Die Lebenslagen älterer Menschen im Spannungsfeld ‚später Freiheit' und ‚sozialer Disziplinierung': forschungsleitende Fragestellungen. In G. M. Backes, & W. Clemens (Hrsg.) Lebenslagen im Alter. Gesellschaftliche Bedingungen und Grenzen (S. 31-52), Opladen: Leske + Budrich, 2000.

[Sø12] Sørensen, E.: STS und Politik. In St. Beck, J. Niewöhner, & E. Sørensen (Hrsg.) Science and Technology Studies. Eine sozialanthropologische Einführung (S. 191-221). Bielefeld: transcript, 2012.

[SSW10] Struve, D., Sengpiel, M., & Wandke, H.: Adaptive Lernunterstützung zur interaktiven Systemnutzung für ältere Benutzer (ALISA). Berlin: GRIN Verlag, 2010.

[Wa02] Wajcman, J.: Gender in der Technologieforschung. In U. Pasero, & A. Gottburgsen (Hrsg.) Wie natürlich ist Geschlecht? Gender und die Konstruktion von Natur und Technik (S. 270-289). Wiesbaden: Westdeutscher Verlag, 2002.

[Wo91] Woolgar, S.: Configuring the User: The Case of Usability Trials. In J. Law (Hrsg.) A Sociology of Monsters: Essays on Power, Technology and Domination. London: Routledge, 1991.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

7. Workshop zur Deklarativen Modellierung und effizienten Optimierung komplexer Probleme (MOC 2015)

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

MOC 2015 - 7. Workshop zur Deklarativen Modellierung und effizienten Optimierung komplexer Probleme [1]

Petra Hofstedt[2], Ulrich John[3], Armin Wolf[4]

Abstract: Die aktuellen Entwicklungen und Herausforderungen in Gesellschaft und Wirtschaft stellen neue Anforderungen an die Informatik und Wirtschaftsinformatik. Nur durch die Weiterentwicklung und den gezielten Einsatz von IT-Technologien, -verfahren und -methoden schaffen wir es, uns den Herausforderungen unserer Zeit zu stellen und nachhaltige, wirtschaftliche Lösungen zu erarbeiten. Beispieldomänen hierfür sind die Energiewende, das flexible Steuern und Überwachen großer Verkehrs- und -versorgungsnetze, die Digitalisierung und Agilisierung von Wirtschaftsunternehmen und das Schaffen und Betreiben agiler effizienter Unternehmensverbünde. Wesentliche Aspekte hierfür sind „agilitätsbegünstigende" und validierbare Modellierung und die Optimierung von Prozessen und einer effizienten und nachhaltigen Ressourcennutzung. Eine der großen Herausforderungen besteht darin, dass die zu behandelnden Aufgabenstellungen aufgrund Ihrer wachsenden Größe und steigenden Komplexität immer schwerer beherrschbar sind. Gleichzeitig erfordern viele Anwendungsfälle – z.B. in der Logistik und beim Infrastrukturmanagement – gute oder sogar beste Lösungen innerhalb kürzester Zeit und auch deren interaktive Adaption, um z.B. Betriebsabläufe oder Pläne bei Störungen oder Ausfällen unmittelbar, bequem und optimal anpassen zu können. Der Workshop adressiert daher Aspekte der Modellierung, der Optimierung und der Simulation hochkomplexer Systeme im o.g. Sinne. Besonderes Interesse besteht dabei an Verfahren, die neben Effizienz auch Deklarativität und Interaktivität bei der Behandlung von Simulations- und Optimierungsproblemen sichern.
Die zehn Papiere des Workshops wurden vier Sessions zugeordnet, wobei erstmalig eine Studierenden-Session und eine Brainstorming-Session geplant wurden. Neun Arbeiten sind begutachtete Papers, das Kurzpaper „Deklarative Modellierung und effiziente Optimierung - quo vadis? Ausgewählte Herausforderungen und Potenziale für die nächsten Jahre – zur Diskussion" enthält ausgewählte Themenvorschläge mit jeweiliger kurzer Beschreibung für die Brainstorming-Session, die dort und darüber hinaus eine Diskussion im o.g. Sinne anregen sollen. Die für die Studierenden-Session akzeptierten Arbeiten sind „Constrained Production Schedule Optimization of Output-Normalized Expenditure under Uncertainty in Shift Duration and Energy Price Forecasts", "Eine grafische, interaktive Benutzeroberfläche für den Constraint-basierten Produktkonfigurator FdConfig" und „Constraint-basierte Planung und Optimierung von Prüfungsterminen mithilfe einer graphischen Benutzeroberfläche".
Generell soll der Workshop Fachleuten, Anwendern und Interessierten die Möglichkeit zum Austausch und zur nutzbringenden Diskussion von Ideen, Ansätzen, Verfahren und Problemlösungen geben, wobei hierbei idealer Weise Grundlagen für zukünftige, im o.g. Sinne zielführende, Forschungsarbeiten entstehen sollen.

[1] www.digitales-unternehmen.de/MOC2015/MOC2015.html

[2] Brandenburgische Technische Universität Cottbus-Senftenberg, Lehrstuhl Programmiersprachen und Compilerbau, Postfach 101344, 03013 Cottbus, hofstedt@b-tu.de

[3] Hochschule für Wirtschaft, Technik und Kultur (HWTK), Fachbereich Wirtschaftsinformatik, Friedrichstr. 189, 10117 Berlin, ulrich.john@hwtk.de/ john@sir-john.de

[4] Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin, armin.wolf@fokus.fraunhofer.de

Organisation: Petra Hofstedt (Brandenburgische Technische Universität Cottbus - Senftenberg), Ulrich John (Hochschule für Wirtschaft, Technik und Kultur, Berlin), Armin Wolf (Fraunhofer FOKUS, Berlin)

Programmkomitee: Christoph Beierle (FernUni Hagen), Ulrich Geske (Universität Potsdam), Hans-Joachim Goltz (Fraunhofer FOKUS, Berlin), Petra Hofstedt (Brandenburgische Technische Universität Cottbus - Senftenberg), Walter Hower (Hochschule Albstadt-Sigmaringen), Ulrich John (Hochschule für Wirtschaft, Technik und Kultur, Berlin), Andreas Kmoch (Toll Collect), Georg Ringwelski (Hochschule Zittau/Görlitz), Hans Schlenker (IBM Deutschland), Andreas Schutt (NICTA & University of Melbourne, Australien), Dietmar Seipel (Universität Würzburg), Armin Wolf (Fraunhofer FOKUS, Berlin)

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Divide & Conquer – Komplexe Constraint-Probleme durch Aufteilen lösen

Peter Sauer[1] Denny Schneeweiß[1], Petra Hofstedt[1]

Abstract: Die Constraint-Programmierung bietet uns hervorragende Möglichkeiten um komplizierte Probleme zu modellieren und zu lösen. Manchmal sind die zu lösenden Probleme jedoch zu komplex, sodass sie mit den üblichen Mitteln der Constraint-Programmierung nicht in annehmbarer Zeit gelöst werden können. Eine Möglichkeit den Lösungsprozess für solche Probleme zu beschleunigen ist die Aufteilung in mehrere kleinere Teilprobleme. Nicht alle Probleme sind gleich gut geeignet durch Aufteilen schneller gelöst zu werden. Einige erfordern einen erheblichen Mehraufwand in der Modellierung, bei anderen ergibt sich eine geeignete Aufteilung nahezu von selbst. Anhand von zwei Beispielen aus der Praxis beschreiben wir mögliche Strategien um komplexe Constraint-Probleme durch Aufteilen schneller zu lösen.

Keywords: Constraint-Programmierung, CSP

1 Einleitung

In der Constraint-Programmierung werden Probleme mit Hilfe von Bedingungen (Constraints) modelliert und anschließend von einem geeigneten Constraint-Solver gelöst. Hierbei stehen dem Constraint-Solver neben allgemeinen Such- und Lösungsstrategien zur Beschleunigung des Lösungsprozesses zusätzlich sogenannte globale Constraints mit spezialisierten Algorithmen zu Verfügung. Da Constraint-Probleme jedoch sehr komplex sein können, ist die Berechnung von Lösungen in adäquater Zeit jedoch auch dann nicht gesichert. Eine weitere Möglichkeit den Lösungprozess zu beschleunigen, ist die Aufteilung des Constraint-Problems in Teilprobleme welche potentiell auch parallel behandelt werden können und die anschließende Kombination der Teillösungen. Dieses Vorgehen führt i.d.R. jedoch zu weiteren Constraints und weiterem Aufwand, um die Konsistenz der entstehenden Teillösungen garantieren zu können.

Unser Ziel ist es, geeignete Aufteilungen für schwere Constraint-Probleme zu finden, um eine Lösung in adequater Zeit bestimmen zu können.

Wir betrachten im Folgenden die Behandlung und Lösung komplexer Planungsprobleme durch ihre Aufteilung in Teilprobleme und deren Lösung. In Abschnitt 2 stellen wir für zwei Planungsprobleme jeweils deren Struktur, Umfang und mögliche Aufteilungen vor: Abschnitt 2.1 diskutiert die Planung von Prüfungsterminen an einer Universität, Abschnitt 2.2 skizziert ein Dienstplanungsproblem.

[1] Brandenburgische Technische Universität Cottbus – Senftenberg, Lehrstuhl Programmiersprachen und Compilerbau, Postfach 101344, 03013 Cottbus, {peter.sauer, schneden, petra.hofstedt}@b-tu.de

Da solche, aus der Aufteilung eines Planungsproblems resultierende, Teilprobleme i.d.R. nicht voneinander unabhängig sondern globalen Bedingungen unterworfen sind, ist eine nebenläufige oder parallele Behandlung und Lösung zwar möglich, aber mit weiteren Bedingungen und u.U. erheblichem Koordinationsaufwand zur Kombination der Teillösungen verbunden. Berechnet man beispielsweise Teildienstpläne für Zeiträume oder Mitarbeitergruppen, so müssen diese aufeinander abgestimmt werden. Mehrere separat arbeitende Constraint-Solver müssen also Teil- und Zwischenergebnisse austauschen und auf Widersprüche und Konflikte reagieren können.

Haben wir eine Aufteilung eignet sich als Framework zur Formalisierung und Lösung solcher Constraint-Probleme die *Verteilte Constraint-Programmierung* [YH00, Sa12]. Verteilte Constraint-Speicher halten dabei die Variablen und Constraints eines Teilproblems und errechnen hierfür zulässige (Teil-)Lösungen. Abhängigkeiten zwischen Teilproblemen werden in einer Hierarchie der Speicher abgebildet. Teil- und Zwischenergebnisse sowie Konflikte werden in Form von Constraints zwischen den Speichern entsprechend der Hierarchie ausgetauscht. In Abschnitt 3.1 präsentieren wir kurz verteilte und nebenläufige Constraint-Speicher. Anschließend illustrieren wir in Abschnitt 3.2 ihre Verwendung zur Lösung komplexer Constraint-Probleme durch Aufteilung.

2 Aufteilung in Teilprobleme

Nicht alle Probleme eignen sich gleich gut, um durch Aufteilen schneller gelöst zu werden. Bei der Planung von Prüfungsterminen an einer Universität liegt eine geeignete Aufteilung in Teilprobleme sehr nahe. Im Gegensatz dazu lässt sich für die Erstellung von Rahmendienstplänen so leicht keine geeignete Aufteilung in Teilprobleme finden. In den beiden folgenden Abschnitten spezifizieren wir die beiden genannten Probleme und deren Anforderungen genauer und stellen jeweils Strategien zu deren Aufteilung in mehrere Teilprobleme vor.

2.1 Planung von Prüfungsterminen an einer Universität

Das Problem der Prüfungsplanung an der BTU Cottbus – Senftenberg (am Zentralcampus) ist von vornherein hierarchisch strukturiert, so dass es sich für eine nebenläufige oder parallele Lösung hervorragend eignet. Im Folgenden stellen wir das Problem, seine Struktur und den Lösungsprozess detaillierter dar.

Problemstellung: Das hier beschriebene Problem der Prüfungsplanung umfasst die in Cottbus zentral geplanten (schriftlichen) Prüfungen. Dies betrifft insbesondere Prüfungen der Module der Bachelor-Studiengänge, da die Studienpläne in den Masterstudiengänge i.d.R. individuell von den Studierenden (in Absprache mit ihrem Mentor) zusammengestellt werden. Nicht berücksichtigt werden also Prüfungen in den Masterstudiengängen, die mündlich durchgeführt werden und deren Termine von den Prüfern oder Planern in einem zweiten Planungsschritt nach der zentralen Prüfungsplanung festgelegt werden.

Die zentrale Prüfungsplanung umfasst ca. 500 Prüfungen über 36 Studiengänge je Semester. Der verfügbare Planungszeitraum gliedert sich in zwei zweiwöchige Zeiträume (im Folgenden Zeitraum A und B genannt), die am Anfang bzw. Ende der vorlesungsfreien Zeit bzw. der sogenanten Semesterpause liegen. Es stehen 60 Räume bzw. Hörsäle unterschiedlicher Größen zwischen 10 und 500 Plätzen zur Verfügung.

Für die Festlegung der Prüfungstermine gibt es diverse Nebenbedingungen und Wünsche, die bei der Planung zu berücksichtigen sind. Dies umfasst nicht nur notwendige Raumgrößen oder die Prüfungsdauer (zzgl. Vor- und Nachbereitungszeiten), sondern auch potentielle Kollisionen von Prüfungen innerhalb von Studiengängen, die Einplanung von freien Tagen zwischen Prüfungen eines Studiengangs bis hin zu individuellen Wünschen einzelner Prüfer.

Strukturierung des Lösungsprozesses: Eine Aufteilung des Lösungsprozesses des Planungsproblems in drei Phasen bietet sich an:

1. Zuordnung der Prüfungen auf konkrete Prüfungstage („Tagesplanung“),
2. Zuordnung der Prüfungen auf konkrete Zeitblöcke der im vorherigen Schritt fixierten Tage („Blockplanung“) und
3. Zuordnung von Räumen, wobei Tag und Block der Prüfungen schon feststehen („Raumplanung“).

Eine Zerlegung des Problems in dieser Form ist notwendig, da das Gesamtproblem sehr komplex ist und die Lösung damit in vertretbarer Zeit nicht garantiert werden kann. Auch unterstützt eine solche Aufteilung die Planung von Problemen gleicher Struktur mit deutlich größeren Umfang.

Gleichzeitig ist zu berücksichtigen, dass eine solche Problemzerlegung i.d.R. zu einer Backtracking-artigen Berechnung führen kann. Das heißt, eine bereits erzeugte Lösung für ein Teilproblem der Ebene i (z.B. Ebene $i = 1$, d.h. Tagesplanung) kann in höheren Ebenen $j > i$ (z.B. $i = 2$, d.h. Zeitplanung) zu einem Konflikt führen, so dass auf Ebene i eine neue Lösung erzeugt werden muss. Sind die Bedingungen im Planungsproblem schwach und die Domänen ausreichend groß, können durch die Aufteilung des Lösungsprozesses Lösungen u.U. ohne Backtracking erzeugt werden.

Constraints für unser Problem (nachfolgend detaillierter beschrieben) können jeweils eindeutig einer Planungsphase zugeordnet werden. Das heißt, die Zerlegung des Problems in Phasen impliziert keine zusätzlichen Constraints. Mit dieser Aufteilung gehen durch den phasenweisen Lösungsprozess auch keine Lösungen verloren. Eine solche Aufteilung kann allerdings nicht für alle Probleme gefunden werden, was wir in Abschnitt 3.2 diskutieren.

Wir diskutieren nun die einzelnen Lösungsphasen und damit verbundenen Constraint-Probleme.

(1) Tagesplanung

Die Tagesplanung regelt die Zuordnung aller Prüfungen auf konkrete Tage im Planungszeitraum. Wichtige komplexe Constraints auf dieser Ebene sind durch die Zuordnung von Prüfungen zu Fachsemestern der einzelnen Studiengängen gegeben: Da es sich bei den zu planenden Prüfungen um Bachelor- und damit häufig um Grundstudiumsmodule (wie Mathematik-, BWL- oder Programmierungsgrundlagen) handelt, sind diese oft mehreren, teilweise auch sehr vielen Studiengängen gleichzeitig zugeordnet.

Wir berücksichtigen folgende Constraints:

- Für jedes Fachsemester eines Studiengangs (im Folgenden "FS-Studiengang") darf höchstens eine Prüfung pro Tag stattfinden. Variante: Es muss mindestens ein (oder x) freie Tage zwischen zwei Prüfungen eines FS-Studiengangs liegen.
- Prüfungen eines FS-Studiengangs dürfen nicht mit Prüfungen des gleichen Studiengangs des vorangehenden Studienjahrs kollidieren (Berücksichtigung von Studierenden, die Prüfungen wiederholen müssen).
- Nur Prüfungen mit dem Wunsch 'Sonnabend', sollen auch am Sonnabend geschrieben werden. Prüfungen mit dem Wunsch Zeitraum A oder Zeitraum B (s. oben) sollen auch in diesem Zeitraum stattfinden.

Einige dieser Constraints können in Varianten des Problems auch als Optimierungskriterium eingesetzt werden. Beispielsweise kann die Maximierung der Anzahl freier Tage zwischen Prüfungen eines FS-Studiengangs angestrebt werden.

Eine weitgehend gleichmäßige Verteilung aller Prüfungen über den gesamten Zeitraum erleichtert die Lösungsfindung in den Ebenen (2) und (3) (Block- und Raumplanung).

(2) Blockplanung

Auf der Ebene der Blockplanung gehen wir davon aus, dass die Prüfungen (in Ebene (1), d.h. Tagesplanung) bereits konkreten Tagen zugeordnet wurden. Nun werden konkrete Zeitblöcke fixiert.

Jeder Prüfungszeitraum, d.h. A und B, umfasst 12 Tage (2 Wochen ohne Sonntag), täglich jeweils von 7:30 bis 19:00 Uhr. Übliche Prüfungsdauern liegen zwischen 60 und 180 Minuten, wobei Zeiten der Vor- und Nachbereitung hinzugerechnet werden müssen. Jeder Prüfungstag wird in 23 Böcke zu je 30 Minuten aufgeteilt, um eine gute Ausnutzung von Zeiten und Räumen gewährleisten zu können.

Da die Prüfungen bereits konkreten Tagen zugeordnet wurden, sichern die Constraints dieser Ebene vor allem die Bereitstellung von hinreichend langen, zusammenhängenden Zeitblöcken für jede Prüfung ab.

Auch für Ebene (2) Blockplanung gilt: Eine weitgehend gleichmässige Verteilung aller Prüfungen über den betreffenden Prüfungstag erleichtert die Lösungsfindung bzw. Planung

auf der nächsten Ebene (3) Raumplanung. Problemvarianten erlauben hierbei aber noch eine Bevorzugung bestimmter Blöcke, so wird — soweit möglich — der erste und letzte Prüfungsblock vermieden.

(3) Raumplanung

Die Raumplanung geht von einer konsistenten Zuordnung von Tagen und Zeiträumen je Prüfung aus und ordnet den Prüfungen Räume zu.

Constraints dieser Ebene berücksichtigen folgende Bedingungen:

- Die Summe der Plätze soll mindestens der der erwarteten Teilnehmer entsprechen.
- Gleichzeitig soll die Platzanzahl die Teilnehmerzahl nicht um Größenordnungen übertreffen.
- Die Einplanung mehrerer Räume für eine Prüfung wird unterstützt (notwendig bei Prüfungen mit besonders hohen Teilnehmerzahlen). Werden mehrere Räume für eine Prüfung gleichzeitig gebucht, so gilt:
  - Hörsäle sollen nicht mit kleinen Räumen kombiniert werden.
  - Es sollen so wenige Räume wie möglich pro Prüfung gewählt werden.
- Räume dürfen nicht zu einer Zeit mehrfach verplant werden.

Wurden in der Raumplanung für die gegebene Zuordnung von Tagen und Zeiträumen passende Räume gefunden, so haben wir eine Gesamtlösung gefunden, welche dem Nutzer präsentiert werden kann.

2.2 Dienstplanung einer Feuerwehrleitstelle

Problemstellung: Dieses Problem umfasst die automatische Rahmendienstplanung im Schichtbetrieb am Beispiel der *Regionalleistelle Berlin-Brandenburg* in Cottbus, die für das gesamte Gebiet Südbrandenburgs und den Flughafen Berlin-Brandenburg International mit insgesamt ca. 650.000 Einwohnern zuständig ist.

Ein Rahmendienstplan wird viele Monate im Voraus erstellt und ordnet die Mitarbeiter tageweise in verschiedene Schichten ein, wobei planbare Fehlzeiten wie Urlaub oder Weiterbildungen noch nicht berücksichtigt werden. Dies geschieht später im Grunddienstplan, für den der Rahmendienstplan die Basis bildet. Kurzfristige Änderungen durch krankheitsbedingte Fehlzeiten oder Schichttausch werden in der Tagesdienstplanung berücksichtigt, welche den letzten Schritt der Dienstplanung bildet. An dieser Stelle soll nur die Rahmendienstplanung betrachtet werden, da diese langfristig erstellt werden muss und durch den großen Betrachtungszeitraum und die Vielzahl von Constraints eine hohe Komplexität besitzt. Die Erstellung der weiterführenden Grund- bzw. Tagespläne hat dann eine wesentlich geringere Komplexität, da die Erfüllung der wichtigsten Rahmenbedingungen bereits vom Rahmendienstplan gewährleistet wird.

Planungsanforderungen

Für die Rahmendienstplanung müssen eine ganze Reihe organisatorischer Anforderungen der Leitstelle wie auch arbeitsrechtliche Rahmenbedingungen und arbeitswissenschaftliche Erkenntnisse beachtet werden, die wir im Anschluss näher erläutern werden. Ziel der Dienstplanung ist die Sicherstellung der vollen Einsatzbereitschaft der Leitstelle durch notwendige Personalstärken in den verschiedenen Schichten. Um diese auch langfristig zu garantieren, muss die Dienstplanung jedoch auch auf die physiologischen, sozialen und psychischen Auswirkungen der Arbeitsbelastung im Schichtbetrieb Rücksicht nehmen und diese Faktoren in die Planung mit einfließen lassen.

In der Regionalleistelle existieren insgesamt neun Schichttypen. Neben Frühschicht (6:00-14:00 Uhr), Spätschicht (14:00-22:00 Uhr) und Nachtschicht (22:00-6:00 Uhr) gibt es drei weitere sogenannte Springerschichten die parallel dazu laufen und zur Abdeckung von Fehlzeiten durch Krankheit oder Urlaub dienen. Fällt ein Mitarbeiter in den normalen Schichten aus, kann ein Springer direkt übernehmen. Weiterhin existieren noch zwei Tagesdienstschichten die 8:00-16:00 Uhr bzw. 11:00-19:00 Uhr laufen sowie eine Praktikumsschicht (7:00-18:00 Uhr) für Ausbildungstätigkeiten und das Training von Rettungstätigkeiten.

Constraints: Die grundlegenden Constraints für die Dienstplanung ergeben sich aus einer Bedarfstabelle der Leitstelle (siehe Tabelle 1), in der für jeden Schichttyp und Wochentag die zur Einsatzbereitschaft notwendige Mitarbeiterzahl (Schichtstärke) verzeichnet ist. Die Leitstelle verfügt über ca. 40 Mitarbeiter.

	Mo	Di	Mi	Do	Fr	Sa	So
Frühschicht	7	7	8	8	8	7	7
Spätschicht	6	6	7	6	7	6	6
Nachtschicht	5	5	5	5	5	5	5
Praktikum	1	1	1	1	1	0	0
Tagesdienst 1	1	1	0	0	0	0	0
Tagesdienst 2	0	0	0	0	0	0	0
Springer Früh	2	2	2	2	2	2	2
Springer Spät	2	2	2	2	2	2	2
Sprinter Nacht	2	2	2	2	2	2	2

Tab. 1: Zur Einsatzbereitschaft notwendige Schichtstärken der Regionalleitstelle Lausitz für eine Belegschaft mit 38 Mitarbeitern [Hä13].

Die Dienstplanung muss diese Schichtstärken für die jeweiligen Tage garantieren, wobei eine Reihe weiterer Constraints hinzukommen, die sich aus den gesetzlichen Regelungen zum Arbeitsschutz ergeben:

- Der Gesetzgeber gibt im Arbeitszeitgesetz für Schichtarbeit vor, dass diese acht Stunden in der Regel nicht überschreiten darf.

- Weiterhin müssen Ruhezeiten von mindestens elf Stunden zwischen zwei Schichten eingehalten werden, was Auswirkungen auf die Plangestaltung hat. Auf eine Nachtschicht dürfen daher keine Früh- oder Spätschicht folgen, da hierbei die Ruhezeit unterschritten wird.
- Die Länge von Blöcken aufeinanderfolgender Schichten des gleichen Typs muss mindestens zwei und höchstens vier Tage bzw. Nächte dauern, um eine negative Veränderung des Tag-Nacht-Rhythmus zu verhindern.
- Eine Vorwärtsrotation beim Schichtwechsel von Früh- über Spät- nach Nachtschicht hat sich als vorteilhaft erwiesen [Be05].
- Weiterhin sollen für die Tage am Wochenende bestenfalls freie Tage, ansonsten aber möglichst gleiche Schichttypen geplant werden.
- Außerdem muss der Freizeitausgleich möglichst geblockt erfolgen.
- Nach Schichten an Sonn- und gesetzlichen Feiertagen müssen spätestens zwei bzw. acht Wochen später Ausgleichstage gegeben werden.
- Weiterhin dürfen an den Wochenenden keine Praktika stattfinden.

Lösungsansatz: Die Dienstplanerstellung erfolgt wochenweise und rollierend. Dabei wird ein Dienstplan für einen einzigen Mitarbeiter über eine Anzahl von Wochen erstellt. Für jeden weiteren Mitarbeiter wird diese Einzelplanung dann jeweils um eine Woche verschoben verwendet. Dies garantiert eine faire Verteilung der Arbeitslast auf alle Mitarbeiter. Für die Leitstelle in Cottbus ergibt sich damit ein angestrebter Planungszeitraum von ca. 40 Wochen, welcher allerdings aufgrund der Problemkomplexität nicht in vertretbarer Zeit berechnet werden kann.

Daher muss das Gesamtproblem in kleinere Teile unterteilt werden, die sich in akzeptabler Zeit lösen lassen. Für die Rahmendienstplanung eignet sich die Unterteilung in mehrere Wochenblöcke. Vor allem Constraints, die sich lediglich auf Wochentage beziehen, lassen sich dadurch lokalisieren. Beispiele hierfür sind die Relegung, dass an Wochenenden stets gleiche Schichttypen bzw. Freizeitausgleich geplant werden soll und das dort keine Praktikumsschichten stattfinden.

Für andere Constraints gilt dies leider nicht. Die Anforderungen zur Sicherstellung der Schichtstärken und die meisten gesetzlichen Rahmenbedingungen müssen nicht nur lokal für die einzelnen Problemteile, sondern global über den gesamten Planungszeitraum erfüllt sein. Zum Beispiel darf die mittlere Arbeitsbelastung über den gesamten Planungsbereich täglich acht Stunden nicht überschreiten. Ausgleichstage für Feiertagsarbeit müssen in einem Zeitraum von zwei bzw. acht Wochen, also ggf. über Teilproblemgrenzen hinweg, gewährt werden.

Im ersten Schritt des Lösungsprozesses muss zunächst lokale Konsistenz für die Teilprobleme hergestellt werden, sodass Werte-Belegungen, die zu keiner Lösung im Teilproblem

führen, eliminiert werden. Um das Gesamtproblem zu lösen muss anschließend globale Konsistenz erreicht werden. Dafür sind zwei Verfahren geeignet:

Die erste Variante ist eine Lösungssuche mittels verteiltem Backtracking über den Teilproblemen, wobei die miteinander kommunizierenden Agenten versuchen, sich auf eine Gesamtlösung zu einigen.

Eine weitere Möglichkeit besteht in der Redefinition von Constraints, die sich nicht natürlich lokalisieren lassen, in schärfere Constraints auf lokaler Ebene. Bei einer Aufteilung des Gesamtproblems von vierzig Wochen in vier Blöcke mit je zehn Wochen kann beispielsweise die lokale Schichtstärke auf den Faktor 0.25 bis 0.3 reduziert werden. Durch diese Verschärfung können Gesamtlösungen verloren gehen. Ist der Lösungsraum jedoch ausreichend groß, können genügend Lösungen übrig bleiben, sodass das Verfahren eine effiziente Lösungsstrategie bietet. Erfahrungen aus der Anwendungsdomäne helfen beim Finden geeigneter schärferer Contraints.

3 Verteilte Lösung von CSPs

Einen geeigneten Formalismus in mehrere Teilprobleme aufgeteilte Probleme zu lösen bietet die verteilte Constraint-Programmierung. Diese wird im folgenden Abschnitt näher erläutert. In Abschnitt 3.2 beschreiben wir, wie mit Hilfe der verteilten Constraint-Programmierung die oben beschriebenen Probleme schneller gelöst werden können.

3.1 Verteilte Constraint-Programmierung

Die verteilte Constraint Programmierung [YH00, Fa06] (engl. *Distributed Constraint Programming*) befasst sich mit der nebenläufigen und parallelen Lösung von CSPs (Constraint Satisfaction Problems).

Constraint Satisfaction Problems: Ein *CSP* besteht aus einer Menge $\{x_1, x_2, \ldots, x_n\}$ von n Variablen, welche jeweils Werte aus gegebenen *Domänen* $D_1, D_2, \ldots, D_n$ annehmen können, sowie einer Menge von Constraints über diesen Variablen. Dabei beschreiben die Constraints Bedingungen, welche als Relationen modelliert werden können. Ein Constraint $c(x_{c,1}, \ldots, x_{c,j})$ ist eine Teilmenge aus dem kartesischen Produkt $D_{c,1} \times \ldots \times D_{c,j}$ und beschreibt alle Kombinationen von Belegungen für die Variablen $x_{c,1}, \ldots, x_{c,j}$ unter denen das Constraint c gilt. Eine *Belegung* ist dabei die Zuordnung von Werten $v_k \in D_k$ zu den entsprechenden Variablen x_k. Eine Belegung für die Variablen des CSPs unter der das CSP (d.h. alle Constraints des CSPs) erfüllt ist, wird als *Lösung* des CSP bezeichnet. Ein CSP ist *konsistent*, wenn mindestens eine Lösung existiert, anderenfalls ist es *inkonsistent*.

Das Lösen eines CSP, also die Suche nach einer Belegung für alle Variablen, unter der alle Constraints erfüllt sind, ist im allgemeinen NP-vollständig. Eine Vielzahl von Algorithmen, Techniken und Heuristiken zielen daher auf eine Beschleunigung des Lösungsprozesses, um CSPs handhabbar zu machen ([RvBW06]).

Distributed CSPs: Bei einem *verteilten CSP* (engl. *distributed CSP*) werden die Variablen und Constraints auf mehrere Agenten verteilt. Auf der Basis eines nachrichtenorientierten Kommunikationsmodells versuchen die Agenten gemeinsam konsistente Lösungen für das CSP zu berechnen.

Während in der allgemeinen Definition verteilter CSPs zunächst jedem Agenten auch mehrere Variablen zugeordnet werden, wird dies i.d.R. bei der Betrachtung konkreter Lösungsalgorithmen, wie synchrones oder asynchrones Backtracking, eingeschränkt, sodass jedem Agenten genau eine Variable zugeordnet wird.

Dies impliziert jedoch, dass globale Constraints, welche typischerweise über dedizierte Algorithmen effizient gelöst werden, auf binäre Constraints zwischen den Agenten herrunter gebrochen werden müssen. In unserer Implementierung hält ein Agent mehrere Variablen und wir können globale Constraints auf diesen weiterhin mit ihren dedizierten Algorithmen lokal innerhalb eines Agenten bearbeiten.

Yokoo [YH00] grenzt die „verteilte Constraint-Programmierung“ von den echt parallelen Contraint-Lösungsmethoden („parallel/distributed processing“) ab, deren hauptsächliche Motivation Effizienz ist. Als Motivation für die verteilte Constraint-Programmierung führt er stattdessen an, dass hier bereits eine natürliche Verteilung des Problems oder des Wissens auf verschiedene Agenten a priori vorliegt.

Als Beispiel für Probleme mit inhärenter Verteilung auf Agenten nennt Yokoo insbesondere Multiagenten-Truth-Maintenance-Systeme, wobei die Konsistenz oder Lösungssuche auf mehrstufigen Verhandlungsprotokollen basiert. Es wird auch ein Beispiel eines Dienstplanungsproblems genannt, allerdings besteht dieses ebenfalls von vornherein aus klar abtrennbaren Teilproblemen in Form von weitgehend unabhängig planbaren Dienstplänen unterschiedlicher Abteilungen eines Betriebes.

Verallgemeinerte verteilte CSPs: Wir zielen darauf, Constraint-Probleme hinsichtlich ihrer natürlichen Verteilungsmöglichkeiten zu untersuchen und die Idee der verteilten Constraint-Programmierung gezielt zur Effizienzsteigerung einzusetzen. Sind natürliche „Sollbruchstellen“ in Constraint-Problemen nicht offensichtlich, so können Algorithmen der verteilten Constraint-Programmierung dennoch eingesetzt werden und zu Effizienzgewinnen führen.

Hierfür abstrahieren wir von den von Yokoo konzipierten Agenten und ordnen jedem Constraint-Agenten einen Constraint-Speicher zu. Jeder Agent bzw. Speicher a verwaltet eine Menge von Constraint-Variablen V_a und eine Menge von Constraints C_a. Ein solches Constraint C_a wird daher i.d.R. nicht nur eigene Variablen („interne Variablen“) sondern auch Variablen anderer Speicher („externe Variablen“) umfassen, die dem Agenten bekannt sind, jedoch nicht von ihm verwaltet werden, d.h. zwar gelesen, aber nicht geändert werden können.

Im Unterschied zu Yokoos Agenten besitzt ein Speicher also mehrere Variablen, was uns die schnelle Anwendung globaler Constraints innerhalb eines Speichers erlaubt. Yokoo

selbst diskutiert Agenten mit mehreren Variablen nur kurz und weist darauf hin, dass sein Ansatz (eine Variable pro Agent) keine Einschränkung darstellt. Agenten mit mehreren Variablen können dadurch abgebildet werden, dass mehrere Agenten auf einer Maschine laufen und schnell mit einander kommunizieren können.

Die Agenten stellen für ihre Constraint-Probleme Konsistenz her und erzeugen Teillösungen. Um die durch Abhängigkeiten zwischen den Teilproblemen auftretenden Konflikte zu lösen, folgen die Agenten einer Hierarchie.

In der Hierarchie höher liegende Speicher besitzen ein *Vorschlagsrecht* für Teillösungen, d.h. insbesondere für interne Variablen, die aus Sicht anderer – niedrigliegender Speicher – als externe Variable an deren Teilproblemen beteiligt sind. In der Hierarchie niedriger stehende Speicher haben ein *Vetorecht* und können bei auftretenden Inkonsistenzen die ihnen vorgeschlagene Teillösungen ablehnen. Die Kommunikation in unserem Modell erfolgt ebenfalls nachrichtenbasiert und unser Lösungsverfahren ist stark durch das von Yokoo entwickelte verteilte Backtracking inspiriert. Eine detaillierte Beschreibung des Modells findet sich in [Sa12].

3.2 Zerlegung und Lösung von Constraint-Problemen

Die Aufteilung eines Constraint-Problems in Teilprobleme ist stark von der Anwendungsdomäne abhängig. Mit der Zerlegung ergibt sich teilweise auch direkt eine Hierarchie der Constraint-Agenten/Speicher. Wir skizzieren im Folgenden Aufteilungen für die in Abschnitt 2 beschriebenen Constraint-Probleme. Beide Probleme sind ohne eine Aufteilung in mehrere Teilprobleme nicht in akzeptabler Zeit lösbar.

Prüfungsplanung: Bei der Prüfungsplanung ergibt sich eine erste mögliche Zerlegung direkt aus der Anwendung. Zunächst müssen die einzelnen Prüfungen einem Tag zugeordnet werden. Hierbei sind die in Abschnitt 2.1 beschriebenen Constraints zu beachten. Haben wir eine gültige Zuordnung von Prüfungen zu Tagen, können den Prüfungen Zeitblöcke zugeordnet werden. Mit dem Ergebnis dieser Blockplanung können wir anschließend jeder Prüfung einen oder mehrere geeignete Räume zuordnen.

Es bieten sich also drei Constraint-Speicher an: Jeweils einer für Tages-, Block- und Raumplanung. Die Tagesplanung schlägt ihre Teillösungen der Blockplanung vor, welche wiederum ihre Lösungen an die Raumplanung weiterreicht. Daraus ergibt sich die folgende Hierarchie: Tagesplanung $>$ Blockplanung $>$ Raumplanung.

Die Blockplanung und die Raumplanung besitzen jeweils ein Vetorecht und können bei auftretenden Inkonsistenzen erhaltene Teillösungen ablehnen. Wird eine Teillösung abgelehnt, muss eine neue Lösung für das entsprechende Problem bestimmt werden. Unser Lösungsverfahren entspricht also einem verteilten Backtracking. Hierbei werden jedoch keine einzelnen Variablenbelegungen widerrufen, sondern komplette Lösungen eines Constraint-Problems.

Ein möglicher Ablauf der Lösungsfindung könnte wie in Abbildung 1 dargestellt aussehen. Hier wird zunächst von der Tagesplanung ein Tagesplan (TP 1) errechnet und an die Blockplanung gegeben. Diese errechnet daraus eine Zuordnung von Prüfungen auf Zeitblöcke (BP 1.1) und schlägt diese der Raumplanung vor. Diese versucht zum gegebenen Tages- und Blockplan passende Räume zuzuweisen (RP 1.1.1). Dies schlägt allerdings fehl und so muss eine neue Blockplanung (BP 1.2) errechnet werden, welche erneut der Raumplanung vorgeschlagen wird. Diesmal existiert eine passende Raumbelegung und wir haben eine Lösung für das Gesamtproblem gefunden.

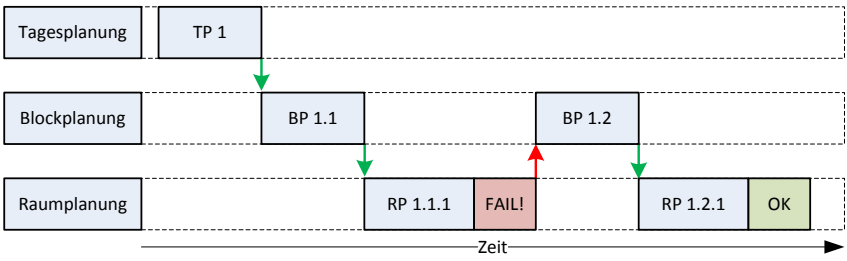


Abb. 1: Eine erfolgreiche Prüfungsplanung mit einem Backtracking-Schritt.

Werden alle Lösung für das Teilproblem des in der Hierarchie am höchsten stehenden Constraint-Speichers abgelehnt, ist das Gesamtproblem unter dieser Aufteilung nicht lösbar.

Haben wir eine Zuordnung von Prüfungen zu Tagen, können die Block- und Raumplanung für jeden Tag unabhängig voneinander erfolgen. Anstelle jeweils eines Speichers für die Block- und Raumplanung könnten wir für jeden Prüfungstag eine Block- und eine Raumplanung einsetzen. In Abbildung 2 ist dies schematisch dargestellt. Dies ermöglicht nach dem Errechnen einer Tageszuordnung die parallele Planung der Zeitblöcke und Räume für jeden Prüfungstag. Sobald jedoch eine einzige Blockplanung den aktuellen Vorschlag der Tagesplanung ablehnt, müssen alle Blockplanungen eventuell bereits errechnete Teillösungen verwerfen und mit dem neuen Tagesplan von vorne anfangen.

Dienstplanung: Bei dem Problem der Dienstplanung ergibt sich aus der Anwendung keine geeignete Aufteilung. Es lässt sich keine Aufteilung des Problems finden, bei der alle Constraints genau einem Teilproblem zugeordnet werden können. Dennoch können wir dieses Problem durch Aufteilen schneller lösen. Hierzu gibt es zwei Strategien: die Aufteilung von Constraints und das Hinzufügen neuer Constraints zur Wahrung der Konsistenz.

Wir wollen das Gesamtproblem lösen, indem wir anstelle eines großen Rahmendienstplanes über viele Wochen mehrere Teilrahmendienstpläne über wenige Wochen erstellen. So könnte ein Rahmendienstplan über 32 Wochen in zwei Zehnwochenpläne und einen Zwölfwochenplan aufgeteilt werden. Im Gegensatz zur Prüfungsplanung sind diese Teil-

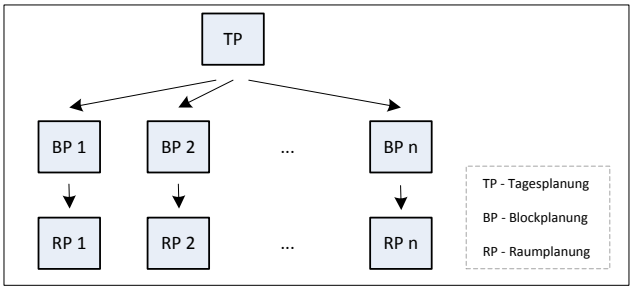


Abb. 2: Block- und Raumplanung können für jeden Prüfungstag unabhängig voneinder bestimmt werden.

probleme stark voneinander abhängig. Viele Constraints, wie zum Beispiel die Garantie der Schichtstärken, beziehen sich auf die Gesamtheit der zu planenden Wochen.

(1) Aufteilung von Constraints

Eine Möglichkeit, das Problem in Form von Teilproblemen zu lösen, ist die Aufteilung von Constraints. Dadurch gehen uns in der Regel einige Lösungen verloren.

Ein Beispiel: Nehmen wir an, wir wollen einen Rahmendienstplan für 32 Wochen erstellen und haben dieses Problem in die Planung von zwei Zehnwochenplänen und einem Zwölfwochenplan aufgeteilt. Wenn wir für Montag sieben Mitarbeiter in der Frühschicht brauchen, ist das als Constraint über den 32-Wochenplan leicht zu modellieren. In mindestens sieben der 32 Wochen muss für Montag 'Frühschicht' stehen. Hinzu kommen natürlich Constraints um sicherzustellen, dass Schichten als Block auftreten usw.

Wollen wir dieses Constraint auf die drei Teilprobleme aufteilen, muss in Summe über alle drei Teilpläne an mindestens sieben Wochen für Montag der Eintrag 'Frühschicht' stehen. Dies können wir zum Beispiel dadurch sicherstellen, dass in den beiden Zehn-Wochenplänen mindestens zwei mal für Montag 'Frühschicht' eingeplant ist und im Zwölf-Wochenplan drei mal.

Dies schließt allerdings einige gültige Verteilungen der Frühschicht am Montag aus und wir verlieren einige Lösungen. Solange wir jedoch einen gültigen Dienstplan erhalten, können wir dies verschmerzen. Diese Strategie eignet sich also besonders für Probleme mit sehr vielen Lösungen. Die Lösungen werden auch sehr schnell gefunden, da kein Backtracking über Teillösungen notwendig ist.

(2) Hinzufügen neuer Constraints

Eine andere Strategie besteht darin, lediglich die für ein Teilproblem lokalen Constraints von dem jeweiligen Constraint-Speicher verwalten zu lassen. Constraints, welche mehrere oder gar alle Teilprobleme betreffen, werden in einen separaten Constraint-Speicher

ausgelagert. Dies sind dann zum Beispiel Constraints, die Schichtstärken betreffen oder Constraints, welche einen nahtlosen Übergang zwischen den Wochenblöcken garantieren. Bei letzteren handelt es sich also um Constraints, welche ohne das Aufteilen in Teilprobleme gar nicht nötig wären.

Alle Constraint-Speicher werden dann wieder in eine Hierarchie eingeordnet, wobei der Speicher mit den Constraints über mehrere Teilprobleme ganz unten in der Hierarchie einzuordnen ist. Er hat also als letztes ein Vetorecht und kann für die Teillösungen überprüfen, ob alle Constraints, welche mehrere Teilprobleme betreffen, eingehalten sind.

Als Lösungsverfahren kommt hier wieder verteiltes Backtracking zum Einsatz. Ein möglicher Verlauf des Lösungsprozesses ist in Abbildung 3 zu sehen.

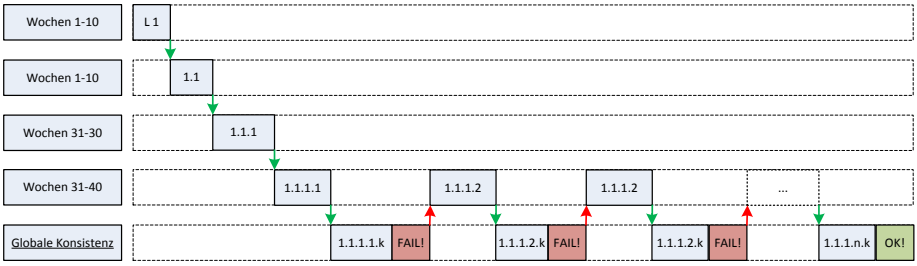


Abb. 3: Verteiltes Backtracking über drei Teilprobleme und einen zusätzlichen Constraint-Speicher, welcher mehrere Teilprobleme betreffende Constraints verwaltet.

Wie in der Abbildung zu erkennen ist, reichen die Speicher ihre Lösungen jeweils an den Speicher zur Bearbeitung des nachfolgenden Wochenblocks weiter. Als letztes erhält der Speicher, welcher mehrere Teilprobleme betreffende Constraints verwaltet, die Lösung und überprüft deren Konsistenz. Dies verursacht natürlich mehrere Backtrackingschritte. Wir verlieren jedoch keine Lösungen.

Durch das Auslagern der Constraints, welche mehrere Teilprobleme betreffen, und das Hinzufügen neuer Constraints, um die Konsistenz der Teillösungen zu garantieren, können wir verhindern, dass uns Lösungen verloren gehen. Der Lösungsprozess dauert aber erheblich länger, da durch die vielen Abhängigkeiten zwischen den Teilproblemen viele Backtrackingschritte notwendig sind. Dennoch kommen wir durch das Aufteilen häufig schneller zur Lösung als durch den Versuch, das Problem im Ganzen zu lösen, da ein Teil der Constraints innerhalb der einzelnen Teilprobleme bearbeitet wird.

4 Zusammenfassung

Anhand der Prüfungsplanung an einer Universität und der Rahmendienstplanung einer Feuerwehrleitstelle haben wir gezeigt, wie Probleme durch Aufteilen schneller gelöst werden können. Beide Probleme eignen sich unterschiedlich gut dazu, durch Aufteilen gelöst zu werden.

Die Planung der Prüfungen ist ein sehr dankbares Problem, da sich eine für den Nutzer intuitive Aufteilung aus der Anwendung ergibt. Die Menge der Constraints kann dabei so aufgeteilt werden, dass jedes Constraint genau einem Teilproblem zugeordnet werden kann. Zusätzlich können die Block- und Raumplanung für jeden Prüfungstag unabhängig voneinander geplant werden. Dies ermöglicht einen hochgradig parallelen Lösungsprozess.

Im Gegensatz dazu ergibt sich bei der Dienstplanung keine intuitive Aufteilung in Teilprobleme. Die Menge der Constraints lässt sich nicht so aufteilen, dass jedes Constraint genau einem Teilproblem zugeordnet ist. Dennoch können wir dieses Problem durch Aufteilen schneller lösen, indem wir bei ausreichend vielen Lösungen auf einige Lösungen verzichten und die Constraints für die Teilprobleme verfeinert lokalisieren. Des Weiteren können Constraints, welche mehrere Teilprobleme betreffen, auch in einen separaten Constraint-Speicher ausgelagert werden und das Problem kann mit Hilfe eines verteilten Backtrackings gelöst werden.

Literaturverzeichnis

[Be05] Beermann, Beate: Leitfaden zur Einführung und Gestaltung von Nacht- und Schichtarbeit. http://www.baua.de/de/Publikationen/Broschueren/A23.html. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), 2005.

[Fa06] Faltings, Boi: Distributed Constraint Programming. S. 699–729. In [RvBW06], S. 699–729, 2006.

[Hä13] Häfner, Franz: Constraint-basierte Berechnung ausgeglichener und korrekter Rahmendienstpläne in akzeptabler Berechnungszeit. Masterarbeit, BTU Cottbus-Senftenberg, 2013.

[RvBW06] Rossi, Francesca; van Beek, Peter; Walsh, Toby: Handbook of Constraint Programming. Elsevier, Amsterdam, The Nethderlands, 2006.

[Sa12] Sauer, Peter: Verteilte Nebenläufige Constraint-Programmierung. Masterarbeit, BTU Cottbus, 2012.

[YH00] Yokoo, Makoto; Hirayama, Katsutoshi: Algorithms for Distributed Constraint Satisfaction. Autonomous Agents and Multi-Agent Systems, 3(2):198–212, 2000.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Über den Umgang mit modellabhängigen Komponenten: Verfeinerungen und ihre Auswirkungen

Michaela Helbig[1], Jens Weller[2], Hajo Wiemer[3] und Gunnar Dietz[4]

Abstract: Die Modellierung von Prozessen spielt in vielen Bereichen eine große Rolle. Dieser Artikel beschäftigt sich damit, was geschieht, wenn ein Modell nicht nur visualisiert, sondern für sich anschließende Aufgaben weiter verwendet wird. In diesem Fall sind weitere Komponenten von dem Modell abhängig und Modellmodifikationen sind nicht mehr trivial. Eben diese Schwierigkeiten treten bei der Verwendung einer im produktionstechnologischen Projektumfeld entstandene Methode auf. Das Beispiel der Verfeinerung von Modellen innerhalb dieser Methode soll zeigen, welche Auswirkungen aus derartigen Abhängigkeiten zu Modellelementen resultieren.

Keywords: Prozessmodellierung, Verfeinerung, modellbasierte Methode

1 Einleitung

In diesem Artikel wird eine Methode aufgegriffen, die im produktionstechnologischen Umfeld für die Entwicklung und Verbesserung von Prozessen entwickelt wurde und deren zentraler Bestandteil in der Prozessmodellierung liegt. Die Modelle bilden die Grundlage für weiterführende Betrachtungen und ganzheitliche Analysen, d.h. sie sind direkt mit anderen Komponenten der Methode verknüpft. Im ersten Teil dieser Arbeit sollen daher grundlegende Erläuterungen über die Ausgangsituation, die Modellierung selbst und ihre Rolle getroffen werden (Kapitel 2). Nach der Erprobung der Methode tritt zunehmend das Bedürfnis nach einer Möglichkeit der hierarchischen Verfeinerung von Prozessen auf. Dieses Beispiel wird im Weiteren (Kapitel 3) dazu verwendet, die Problematik der Modellmodifikation unter Berücksichtigung der modellabhängigen Komponenten zu erörtern. Dieses Kapitel beschäftigt sich zu Beginn mit theoretischen Grundlagen von Modellverfeinerungen und geht danach im Speziellen auf die Gründe, die Auswirkungen und Lösungsansätze sowie die Risiken von derartig gebundenen Modellen ein. Die Arbeit schließt mit einer Zusammenfassung und dem Ausblick auf zukünftige Vorhaben.

[1] TU Dresden, IWM, 01062 Dresden, michaela.helbig@tu-dresden.de
[2] Symate GmbH, Wittenberger Str. 97, 01277 Dresden, jens.weller@symate.de
[3] TU Dresden, IWM, 01062 Dresden, hajo.wiemer@tu-dresden.de
[4] TU Dresden, IWM, 01062 Dresden, gunnar.dietz@tu-dresden,de

2 Motivation und Ausgangssituation

2.1 Voraussetzungen im Projektumfeld

Die Verbesserung und Weiterentwicklung technologischer Prozesse spielt bei der Entwicklung industrieller Produkte eine immer stärkere Rolle. Vor diesem Hintergrund wurden im produktionstechnischen Sektor verschiedene Projekte initiiert, in denen sich jeweils Teilprojekte damit auseinandersetzen, ein datengetriebenes Vorgehen für die effektive Entwicklung und Verbesserung technologischer Prozesse auf Basis von Daten aus und zur Produktion zu erarbeiten. Die entwickelte Methodik ist in Abb. 1 skizziert.

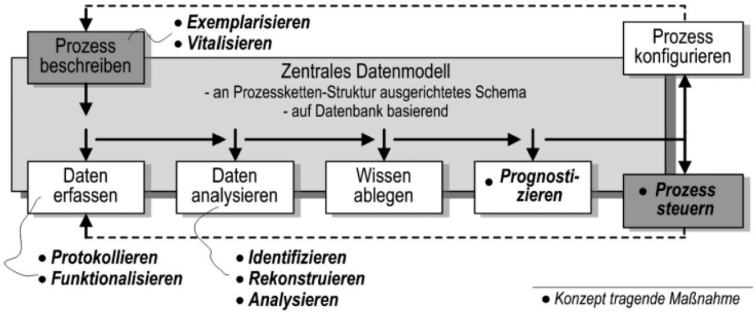


Abb. 1: Methode zur effektiven Entwicklung und Verbesserung produktionstechnischer Prozesse [GW10]

Fertigungsprozesse werden in diesem Ansatz auf operativer Ebene abgebildet. Dazu werden von den am Fertigungsprozess beteiligten Elementen hinsichtlich einer Verbesserungszielstellung relevante Daten berücksichtigt. Dies sind zunächst die festgelegten Bewertungsgrößen am Produkt und schließlich die darauf wirkenden Einflüsse, welche meist Einstellparameter an Maschinen, Eigenschaften der eingehenden Halbzeuge und oft auch Störgrößen u.a. aus der Umgebung sind. Diese Eigenschafts- und Einflussgrößen liegen in den Unternehmen in unterschiedlichen Datenqualitäten vor. Mit Hilfe eines grafischen Modelles wird der Fertigungsprozess mit seinen Elementen und deren Parametern notiert. Darauf aufbauend werden die Daten aus relevanten Quellen wie Messgeräten und Maschinen, Qualitätssicherungssystemen, Lieferscheinen, Anlagen- und Laborbüchern sowie Erfahrungen des Personals eingesammelt, in eine systematische und elektronisch verarbeitbare Form gebracht und in eine zentrale, analysefähige Datenbasis überführt [Wi13].

Wie bereits anhand der kurzen Methodenbeschreibung ersichtlich, spielt die Modellierung eine grundlegende Rolle in diesem Konzept, auf die im Folgenden näher eingegangen werden soll.

2.2 Nutzung von Prozessmodellen

Modellbegriff

Für den Begriff des Modells existiert eine Vielzahl von Definitionen. Eine weit verbreitete Definition ist die nach Schütte [Sc98], nach der ein Modell „das Ergebnis einer Konstruktion eines Modellierers, der für Modellnutzer eine Repräsentation eines Originals zu einer Zeit als relevant mit Hilfe einer Sprache deklariert [...]“ ist. Ebenso kann die Definition nach Stachowiak [St73] herangezogen werden, der über das Abbildungs-, das Verkürzungs- und das pragmatische Merkmal ein Modell eingrenzt. Diese Definitionen sind nicht domänengebunden, das heißt sie sind in der Informatik bzw. Wirtschaftsinformatik genau so zutreffend wie für die Beschreibung produktionstechnologischer Prozesse.

Modellierungssprache und -vorgehen

Die für die Umsetzung des beschriebenen Konzeptes verwendete Modellierungssprache basiert auf dem Input-Throughput-Output-Ansatz. Das heißt, entlang des Materialflusses geht Material mit Anfangseigenschaften in einen Prozess ein, beim Durchlaufen des Prozesses werden die Materialzustände unter Wirkung von Einflüssen der Fertigungsmittel und Umgebung verändert, so dass aus dem Prozess Produkte mit Endeigenschaften austreten. Die entwickelte Modellierungssprache ist eine semi-formale Sprache [Fr99], die auf Basis von [DW97] nach domänenspezifischen Gegebenheiten an die Anforderungen des Konzeptes der modellgestützten Methode angepasst wurde (u.a. in [Gr11]).

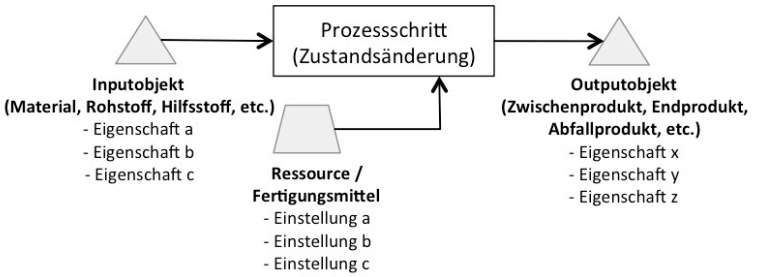


Abb. 2: Elemente der Modellierungssprache (vereinfacht)

Die Grundstruktur der Sprache ist Abb. 2 zu entnehmen. Die Modellierung eines produktionstechnologischen Prozesses läuft damit wie folgt ab:

Während der Modellierung beschreiben die Nutzer den zu untersuchenden Prozess, indem zunächst das Produkt (das Prozessergebnis) mit den zugehörigen Zieleigenschaften, die schließlich der Prüfung und Optimierung unterliegen, betrachtet wird. In den weiteren Modellierungsschritten erfolgt die Abbildung der relevanten Prozessschritte, die den Untersuchungsraum darstellen. Die Untergliederung des

Prozesses erfolgt zweckmäßig entsprechend der Untersuchungsaufgabe. Die Auflösung des Modells kann von einer groben Prozessdarstellung wie einer Lieferkette bis zu einer sehr detaillierten Darstellung wie für einen Operationsablauf in einer Anlage reichen. Zwischen den Prozessschritten werden entsprechend der Sprachdefinition Materialobjekte als Zwischenergebnisse definiert. Wichtig hierfür ist die praktikable Prüfbarkeit der Zwischenergebnisse als Voraussetzung der Datenerfassung und –analyse.

Für jeden Prozessschritt werden die auf die geforderten Eigenschaften des Prozessergebnisses bzw. –zwischenergebnisses wirkenden Einflussparameter definiert. Dieser Modellierungsschritt ist stark erfahrungsgeprägt und erfolgt häufig in einer Gruppe von Fachleuten unterschiedlicher Bereiche, die Schnittstellen zum behandelten Prozess haben. Das können beispielsweise der Technologe, der Maschineneinrichter und –bediener, der Produktionsleiter, der Qualitätsbeauftragte und der Einkäufer sein. Bereits hier wird deutlich, welche unterschiedlichen Sichtweisen auf die produktionstechnologische Prozesskette berücksichtigt werden müssen. Daraus ergeben sich bereits unterschiedliche Fragestellungen, die mit den modellbasierten Untersuchungen behandelt werden müssen.

Rolle der Modellierung im Kontext der verwendeten Methode

Gemäß des Vorgehens innerhalb der zugrundeliegenden Methode, generieren die modellierten Eigenschaften und Einstellungen im nächsten Schritt die Unterpunkte, die Grundlage für die Erfassung in der Datenbank sind. Je nach Datencharakter wird in der Umsetzung eine Erfassungsmaske für die Durchführung des Prozesses generiert oder die (aus bspw. der Maschinensteuerung ausgelesenen) Daten werden direkt in die Datenbank geschrieben und sind somit ebenfalls mit dem Modell verknüpft. Für jede tatsächlich durchgeführte Prozessinstanz können somit Daten erfasst werden und abschließend auf Basis der Modellstruktur über alle Prozessschritte hinweg ausgewertet werden.

Das methodische Konzept basiert (wie bereits in Abb. 1 ersichtlich) auf einem zentralen Datenmodell einer Datenbank, welches mit Hilfe eines Prozessmodells konfiguriert wird und somit den Ausgangspunkt der Generierung und Nutzbarmachung des Prozesswissens darstellt. Somit entstehen Abhängigkeiten zu weiteren Methodenkomponenten (Daten, Analysen, etc.), welche Modellmodifikationen erschweren. Die Zusammenhänge zwischen Modell, Datenbankstruktur und Prozessdaten sind schematisch in Abb. 3 zu sehen.

Die Modellierungssprache bildet den produktionstechnologischen Prozess ab, dient der ingenieurgerechten Beschreibung der prozessrelevanten Zusammenhänge und hilft beim Erzeugen des Prozessverständnisses. Während der experimentellen Untersuchungen oder während der Fertigung werden, zum Zwecke der Protokollierung, Prozessdaten zu den einzelnen Prozessschritten und –zyklen in die durch das Prozessmodell vorgegebene Struktur als Instanzdaten gesammelt.

Die Prozessbeschreibung bildet zudem die Grundlage für weitere methodische Maßnahmen wie Versuchsplanung, Datenanalyse sowie Generierung und Nutzbarmachung von Prozesswissen. Das Prozesswissen wird beispielsweise zur

Prognose und Steuerung des Prozessverhaltens sowie zur Prozesskonfiguration während der Planung angewendet. Methode, Modellierungssprache und Anwendungsbeispiele sind u.a. in [GW10], [Gr10] und [GW11] detailliert beschrieben.

Die Erfassung und Analyse entlang der im Modell festgelegten Prozesskette ermöglicht eine ganzheitliche Prozessbetrachtung über mehrere Prozessschritte und über komplexe Einflussbereiche hinweg. Die ganzheitliche Prozessbetrachtung erlaubt ebenso die Analyse von alternativen Prozessvarianten (sofern diese modelliert wurden), und bietet somit die Grundlage für die Optimierung in der Fertigungsprozessplanung.

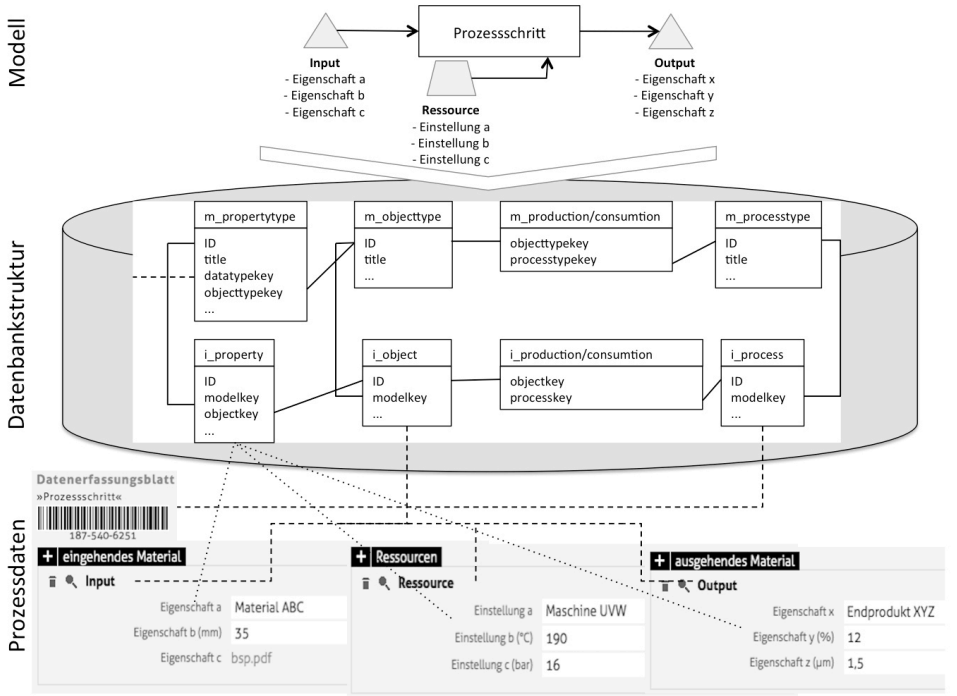


Abb. 3: Abhängigkeiten zwischen Modellkomponenten und Daten (stark vereinfacht)

Die einerseits im Rahmen der Folgenutzung gewinnbringende Verknüpfung der Modelle zu Daten etc., sorgt andererseits für einen erschwerten Umgang mit den Modellen. Aktivitäten im Rahmen der Modellierung sind demnach bei der Verwendung der ganzheitlichen Methodik deutlich weniger trivial als im Bereich der „Nur“-Modellierung. Die beiden oberen Ebenen in Abb. 3 bieten keinerlei Schwierigkeiten im Umgang mit Änderungen. Diese werden durch den Einsatz geeigneter Konzepte von (Modell-) Change bzw. Konfigurationsmanagement [Th07] abgedeckt. Die eigentliche Problematik ergibt sich erst im Umgang mit der Verknüpfung zu den realen Prozessdaten. Ausgehend von direkten Projektanforderungen wurde im Folgenden die Verfeinerung von Prozessmodellen als Beispiel für die Untersuchung zu beachtender

Einflüsse im Umgang mit modellabhängigen Komponenten (in diesem Artikel beschränkt auf Daten) gewählt.

3 Modellverfeinerung im Methodenkontext

Es konnten bereits gute Erfahrungen mit dem oben genannten Ansatz gemacht werden. Zunehmend entstehen jedoch Anforderungen, die Modellierung auf verschiedenen Ebenen durchführen zu können, wofür die derzeitige Modellierungssprache entsprechend erweitert werden soll. Im Folgenden soll daher der Fokus auf der Verfeinerung (top-down) und äquivalent, aber nicht explizit benannt, auf die Vergröberung (bottom-up) von Prozessmodellen und deren Auswirkungen liegen. Hierzu wird ein kurzer Einblick in die theoretischen Grundlagen innerhalb der Modelltheorie gegeben (3.1). Anschließend wird untersucht, wie es innerhalb der Methodennutzung zu möglichen Modellverfeinerungen kommt (3.2), was diese für Auswirkungen haben und wie ggf. darauf reagiert werden kann (3.3) sowie welche Risiken dabei bestehen (3.4).

3.1 Theoretische Grundlagen

In der Modellierung wird das Prinzip der Verfeinerung (ebenso wie der Vergröberung) häufig verwendet. Eine genaue Definition was darunter zu verstehen ist, wird jedoch selten gegeben. Aus Sicht der Wortherkunft bedeutet Verfeinerung etwas feiner, besser und exakter zu machen [Du15]. In anderen Worten heißt das (mit Bezug zur Modellierung), das ein Sachverhalt präzisiert wird.

Betrachtet man exemplarisch Petri-Netze [Ba96] als eine Form der Modellierung, so wird auch da bereits mit den Konzepten der Verfeinerung und Vergröberung gearbeitet. Unter einer Verfeinerung wird dabei eine Konkretisierung der inneren Struktur einer Transition oder Stelle verstanden. Sie ist das Gegenstück zur Vergröberung. Der Einsatz von Verfeinerung und Vergröberung ist abhängig vom gewünschten Abstraktionsgrad und führt zu keiner Änderung im Grundverhalten des Systems.

Neben der Betrachtung von Systemmodellen (Petri-Netz), kann man die Konzepte von Verfeinerung und Vergröberung in ähnlicher Form auch in der Modellierung von (Software-)Systemen finden. Ein Beispiel hierfür ist die UML. In der Spezifikation der Sprache [OM13] wird u.a. bei der Nutzung von Aktivitätsdiagrammen die Funktion „refine“, also verfeinern umgesetzt. Darunter wird eine Beziehung zwischen Modellelementen verschiedener semantischer Ebenen verstanden. Es bestehen „Abstraction Dependencies“ zwischen den einzelnen Modellen. Die Verfeinerung hat auch in diesem Fall keinen direkten Einfluss auf den Modellinhalt, das heißt jedes Modell ist für sich gesehen komplett, beschreibt jedoch den gleichen Prozess wie eine Verfeinerung/Vergröberung.

Die Architektur integrierter Informationssysteme (ARIS) nach Scheer [Sc13], welche

einen Ordnungsrahmen für Geschäftsprozessmodelle unter verschiedenen Sichten liefert, stellt ein weiteres Beispiel dar. Innerhalb der Prozesssicht wird die sogenannte (erweiterte) Ereignisgesteuerte Prozesskette (e)EPK genutzt. Innerhalb der EPK gibt es ebenfalls ein Konzept, welches eine hierarchische Untergliederung von Prozessmodellen erlaubt. Der Prozesspfad (Viereck mit dahinterliegendem Sechseck) steht an der Stelle einer Funktion und repräsentiert einen Wegweiser zu einem detaillierteren Prozessmodell, welches (umklammert durch Prozesspfadelemente) aus dem gleichen Eingangsereignis hervorgeht wie der Prozesspfad und im gleichen Ereignis endet wie das vereinfachte Prozessmodell. Der Prozesspfad bildet somit den Aufruf eines Teils des Hauptmodells, welcher auf übergeordneter Ebene ausgeblendet ist, sowie dem Wegweiser zurück in das Hauptmodell. Der Prozesspfad kann daher genutzt werden, um komplexe Sachverhalte aus dem eigentlichen Prozessmodell auszulagern, aber auch um die Übersichtlichkeit im Modell zu wahren oder mehrere Aktivitäten zu gruppieren.

Zusammenfassend kann gesagt werden, dass es sich bei der Verfeinerung um ein Konzept zum hierarchischen Umgang mit komplexen Aktivitäten handelt, bei dem miteinander verbundene (Teil-)Modelle semantisch gleiche Prozesse auf unterschiedlichen Abstraktionsniveaus darstellen.

3.2 Gründe für Verfeinerungen

Verfeinerungen eines Modells im Kontext der Methode für die effektive Entwicklung und Verbesserung von produktionstechnologischen Prozessen werden i.d.R. aus Darstellungs- oder Übersichtsgründen angelegt. In diesem Fall kann der Prozess beschrieben oder das Modell besser als Kommunikationsgrundlage genutzt werden. Dies ist der einfachste Fall einer Verfeinerung, solange diese nicht zum Zwecke der Datenerhebung genutzt wird. Da die Modellierung innerhalb des methodischen Ansatzes jedoch Basisbestandteil eines ganzheitlichen Konzeptes ist und damit der Annahme unterliegt nicht separat angewendet zu werden, ergeben sich weitere Gründe für die Erstellung von Verfeinerungen daraus, dass weitere Prozessebenen in die Erfassung und Analyse miteinbezogenen werden sollen. Die Nutzung verschiedener Kontextebenen wie in Abb. 4 dargestellt, ist ein Grund für den Bedarf an Modellverfeinerungen, da je nach betrachteter Kontextebene verschiedene Informationen relevant sind oder gar deren Zugang verwehrt werden soll.

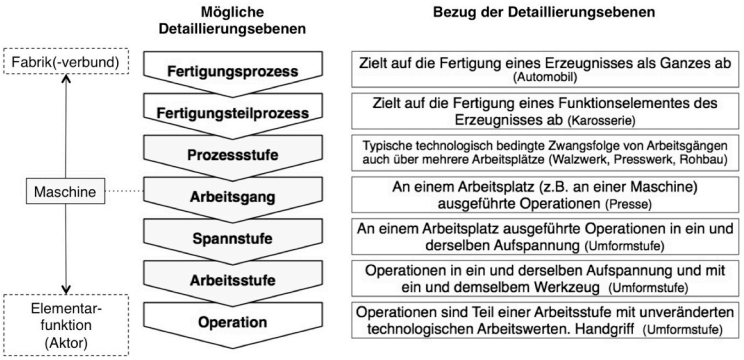


Abb. 4: Kontextebenen aus der Fertigungsprozessgestaltung [JD02]

Zusätzlich zu kontextbasierten Gründen können auch administrative bzw. methodische Gründe anfallen, die eine Verfeinerung der Prozessmodelle notwendig machen. Die folgende Auflistung (basierend auf Projekterfahrungen) fasst die unterschiedlichen Möglichkeiten für Verfeinerungsursachen noch einmal zusammen:

- Übersichtlichkeit, bessere Darstellung
- teilweise unterschiedliche Erfassungsebenen benötigt (z.B. Maschinendaten vs. manuelle Erfassung); Auswertung dann in verschiedenen Ebenen und/oder aggregiert gefordert
- Aufgabenverteilung (jeder sieht nur die für ihn relevanten Ebenen und dazugehörigen Details; aber auch bspw. zum Generieren von Handlungsanweisungen, in denen eine detailliertere Beschreibung notwendig ist)
- Integration von Maschinensteuerung (feingranulare Verfeinerung bis auf Maschinenschrittebene)
- Fabrikplanung (Betrachtung auf höherer Ebene)
- Logistikbetrachtung (Integration von Umgebungsbedingungen und Interaktionseffekten u.a. in Lager- und Transportprozessen)
- Wiederverwendung (z.B. für den Technologietransfer [Wi2013] einzelner Elementarprozesse)

Aus den identifizierten Gründen lässt sich ableiten, welche Modellelemente von einer Verfeinerung betroffen sein können. Die klassische Verfeinerung, bei der Prozesse in feingranulare Prozessschritte mit detaillierteren Eigenschaften zerlegt werden, ist vermutlich der häufigste Anwendungsfall. Darüber hinaus ist es sinnvoll Ressourcen in Baugruppen o.ä. zerlegen zu können oder Umgebungsbedingungen genauer zu detaillieren. Diese Möglichkeiten lassen sich direkt aus der Modellierung ableiten. Auch eine verfeinerte Betrachtung von Eigenschaften (deren Werte bspw. auf höchster Ebene kumuliert werden) ist denkbar. Unter Einbezug der Kontextebenen wird deutlich, das nicht jeder Grund zwingend zu einer Verfeinerung des Prozessmodells (und somit zu

einer anderen Sicht auf das Modell) führt, sondern vielmehr auch zu anderen Lösungen führen kann. Für verschiedene Szenarien ist darüber nachzudenken, die bisherige Modellstruktur aufzuweichen bzw. auf unterschiedlichen Ebenen jeweils andere Modellierungssprachen zu verwenden (z.B. Zustandsautomaten auf Operationsebene). Diese Möglichkeiten sollen im Rahmen dieser Arbeit nicht weiter betrachtet werden. Der Fokus liegt hier auf der Verfeinerung von Modellen innerhalb einer Sprache.

3.3 Auswirkungen und Lösungsvorschlag

Die Verfeinerung bildet demnach eine Sonderform, in der Strukturänderungen durch das Hinzufügen und Ändern von Elementen vorgenommen werden, da hierbei „unechte" Alternativen modelliert werden. Die Verfeinerung auf höchster Ebene verfolgt somit das gleiche Ziel (produktionstechnisch gesehen), wie die auf niedrigster, stellt nur je nach Verfeinerungsstufe einen anderen Detaillierungsgrad des Prozesses dar. Das heißt das es bspw. in der Prozessumsetzung nur eine Möglichkeit (keine „echte" Alternative), aber je nach Verfeinerungsgrad im Prozess mehrere alternative Möglichkeiten („unechte" Modellalternative) für die Erfassung gibt. Wie anhand dieser kurzen Einführung gezeigt, ist eine Modellverfeinerung innerhalb der Methode weniger trivial als eine bloße Verfeinerung der Beschreibung der Abläufe in der (Geschäfts-)Prozessmodellierung. Ein treffenderer Vergleich wäre bspw. die Modifikation von Meta-Modellen. In so einem Fall müssen ebenfalls Konzepte entwickelt werden, wie die darauf basierenden Folgefunktionen zu reagieren haben. Auf den Gründen für Modellverfeinerungen in der Produktionstechnik aufbauend, sind betroffene Elemente identifiziert worden. Die genaue Betrachtung der Konsequenzen führt im Folgenden zu der Ableitung von möglichen Lösungsansätzen.

Wird ein Prozessmodell aus genannten Gründen verfeinert oder vergröbert, können sich folgende Szenarien ergeben:

1. Es sind keine Daten zum Modell vorhanden und eine Erfassung ist nicht geplant.
2. Es sind noch keine Daten vorhanden, aber eine Erfassung ist geplant.
3. Es sind bereits Daten erfasst.

Im ersten Fall handelt es sich um eine unkritische Modellmodifikation, da dieses anscheinend nur zu Beschreibungszwecken genutzt wird. Eine Dokumentation der Veränderung im normalen Modell-KM [Th07] ist ausreichend. Im zweiten und dritten Fall müssen Konzepte für die Erfassung von Daten diskutiert werden. Beim dritten Szenario kommt außerdem hinzu, dass Möglichkeiten der Migration bestehender Daten in Betracht gezogen werden müssen.

Anhand von Abb. 5 soll daher zunächst diskutiert werden, welche zu berücksichtigenden Aspekte aus den dargestellten Verfeinerungsfällen für die Modellierung und den Umgang mit Daten resultieren.

Modellierung

Aus einem Prozessobjekt A werden mehrere Prozessschritte z.B. A.1 und A.2 sowie im Folgenden A.2.1 und A.2.2 aus A.2 werden. Anfang und Ende jeder Verfeinerung bleibt das Objekt was vor der Verfeinerung in den Prozess ein- bzw. aus dem Prozess ausgeht. Für die neuen Prozessschritte werden zusätzliche Materialobjekte (Throughputs) angelegt.

Für den Umgang mit Ressourcenobjekten sind mehrere Möglichkeiten denkbar. Das Ressourcenobjekt wird als Verlinkung wiederholt angebunden. Hierbei wäre zum Zwecke der Übersichtlichkeit das Ein- und Ausblenden der für diesen Prozessschritt relevanten und zu erfassenden Parameter sinnvoll. Zudem kann es sinnvoll sein das Ressourcenobjekt bspw. zur Modellierung einer komplexen Anlage selbst zu verfeinern bspw. in Arbeitsstationen oder Funktionsgruppen z.B. in Teilressourcen 1 und 2 in unterschiedlichen Prozessschritten oder Teilressource 2.1 und 2.2 in einem Prozessschritt.

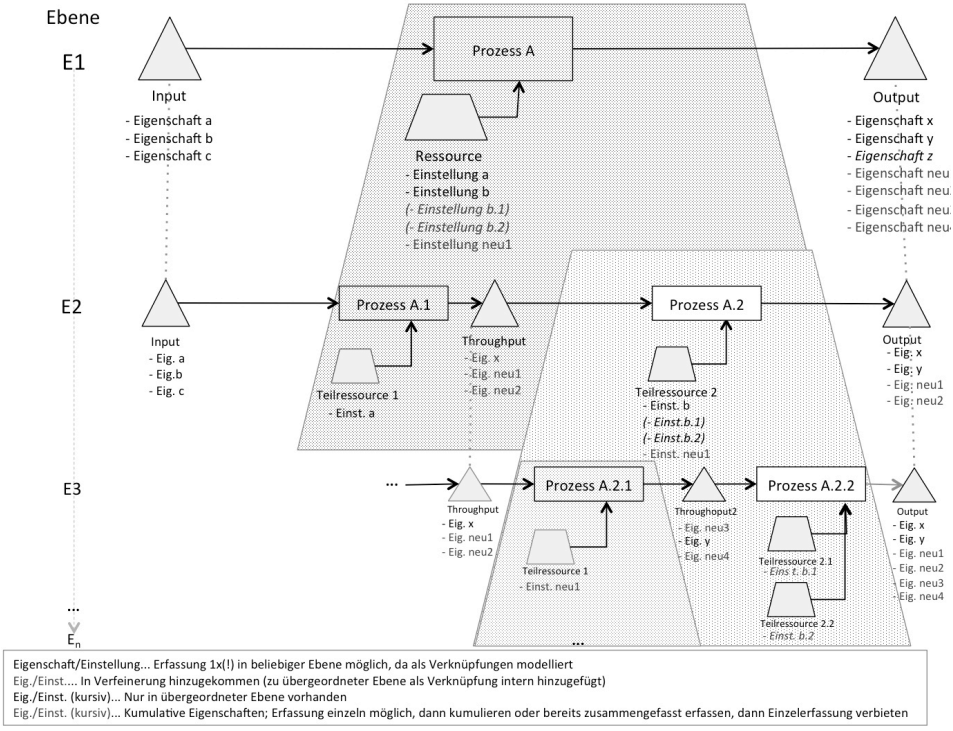


Abb. 5: Verfeinerungsansatz (Beispiel, stark vereinfacht)

Die Objekteigenschaften und Maschineneinstellungen werden je nach Bedarf modelliert. Es ist zu beachten, dass die höheren Ebenen jeweils eine Verknüpfung erhalten, so dass sich folgende Möglichkeiten für die Modellierung von Eigenschaften ergeben:

- Es existieren Eigenschaften, die nur auf oberster Ebene modelliert wurden (Eig. z)
- Die oberste Ebene (E1) hat Kenntnis von allen Eigenschaften des Modells
- Es können Teileigenschaften modelliert werden, die auf höherer Ebene kumuliert werden können (Einstellung b.1 und b.2 entweder separat oder zusammen als Einstellung b oder zu b kumuliert)

Allgemein gilt, das jedes Modellelement genau ein Mal modelliert wird, alles weitere erfolgt durch die Verwendung von Objektreferenzen.

Tabelle 1 fasst die Auswirkungen von Verfeinerungen auf die Modellierung abschließend noch einmal zusammen.

Modellelement	Auswirkungen
a) allgemein	jedes Element nur 1x modelliert, darüber hinaus Verwendung von Referenzen
b) Prozesse	Prozesszerlegung (Bsp.: A→ A.1+A.2) mit jeweils gleichem In- und Output und Erzeugung von Throughputs kann zu c) und d) führen
c) Ressourcen	Wiederverwendung bereits modellierter Ressourcen Zerlegung (Bsp.: Ressource → Teilressource 1+2) ggf. Zuordnung bereits modellierter Einstellungen zu Teilressourcen
d) Eigenschaften und Einstellungen	neue Eigenschaftsmodellierung auf Ebene je nach Bedarf Wiederverwendung mit Referenzen Zerlegung und Kumulation von Eigenschaften (Bsp.: b→b.1 + b.2)

Tab. 1: Zusammenfassung der Auswirkung von Verfeinerungen auf die Modellierung

Umgang mit Daten

Für die Erfassung neuer Daten gibt es zwei Möglichkeiten. Bei der ersten Möglichkeit erfolgt die Erfassung grundsätzlich auf einer festgelegten Ebene z.B. auf der höchsten, niedrigsten oder frei festzulegenden. Bei einer Festlegung ergeben sich Probleme mit der Flexibilität des Systems. Höchste und niedrigste Ebene können sich ändern, sobald darüber oder darunter ein neues Modell erstellt wird. Auch beim Festlegen einer Ebene stellt sich die Frage, wer legt das nach welchen Regeln fest und wie kann mit möglichen Änderungswünschen über die Zeit umgegangen werden. Um diesen Ansatz nutzen zu können, muss die Festlegung fix bleiben und kann auch im Nachhinein nicht geändert werden.

Anhand des Beispiels in Abb. 5 könnte eine Erfassung auf E3 (vorerst detaillierteste Ebene) stattfinden. Dabei kann es dazu führen, dass relevante Eigenschaften, die ursprünglich auf E1 modelliert (Eig. z) und in keiner anderen Ebene referenziert wurden, verloren gehen. Auch die Nutzung von kumulierten Eigenschaften (Einstellung b) ist in

diesem Fall nicht möglich, da nur die Einzelwerte bekannt sind, jedoch nicht ihr Zusammenhang.

Bei einer Erfassung auf höchster Ebene (E1) besteht die Gefahr der Unübersichtlichkeit, da diese Ebene Information über sämtliche Eigenschaften der Verfeinerungsebenen besitzt. Besitzen Teilressourcen beispielsweise die Eigenschaft „Bezeichnung“, so werden bei der Erfassung ggf. mehrere Bezeichnungen aufgelistet, die zwar intern verknüpft, aber augenscheinlich zusammenhanglos aufgezählt erscheinen. Des Weiteren ist beim Umgang mit kumulativen Eigenschaften eine Entscheidung zu treffen, ob diese einzeln erfasst und kumuliert oder als ein Parameter erfasst werden sollen. Die jeweils nicht gewählte Möglichkeit sollte gesperrt werden, da mögliche Tippfehler etc. dazu führen das kumulierter Wert (z.B. b.1 + b.2 könnten Zeiten einzelner Prozessabläufe sein) und einheitlicher Parameter (b könnte Gesamtzeit auf einer Maschine sein) abweichen, was zu erheblichen Problemen bei einer Umsetzung in einem Tool führen kann.

Wählt man eine beliebige Ebene (bspw. E2) für die Erfassung, so ist dies nur sinnvoll, wenn dies explizit von der Projektleitung vorgegeben ist, da eine nicht-administrative Begründung der Ebenenwahl kaum möglich ist. Diese Form der Erfassung kombiniert zudem die Nachteile der Erfassung auf unterster und der Erfassung auf oberster Ebene und sollte daher nicht gewählt werden.

Die zweite Möglichkeit des Erfassens sieht vor das gemischt auf unterschiedlichen Ebenen erfasst werden kann. Dies käme im aktuellen Stand der Methodenentwicklung dem Erfassen in verschiedenen, eigenständigen Modellen gleich. So könnten bspw. im ungünstigsten Fall alle Inputeigenschaften auf Ebene E1, alle Throughput-Eigenschaften und Maschineneinstellung auf Ebene E2 und alle Outputeigenschaften auf Ebene E3 erfasst werden. Außer den bereits genannten Nachteilen würde dieses Erfassungskonzept dazu führen, dass eine Auswertung unmöglich ist. Die Prozessdaten sind mit unterschiedlichen Modellen verknüpft, da sie ja keine Alternativen darstellen, sondern den gleichen Prozess beschreiben. Die durchgängige Betrachtung wäre zerstört.

Damit ist weder eine Erfassung auf einer Ebene, noch die losgelöste Erfassung auf verschiedenen Ebenen geeignet.

Ein Konzept zur Nutzung von Verfeinerungen, welches die modellbasierte Methode ergänzt, muss daher auf die Anforderung eingehen, Daten in beliebigen Ebenen zu erfassen und dennoch eine Prozessauswertung unabhängig der Ebenenstruktur realisieren zu können.

Zusätzlich zu den Verfeinerungen im Modell müssen demnach interne Verknüpfungen auf Seiten der Datenbank vorgenommen werden. Für den direkten Fall heißt das u.a., dass die Teilressourcen direkt zur Ressource auf E1 zuordenbar sind. Die Ressource wird somit zusätzlich mit den Einstellungen verknüpft, die aus der Modellierung der Verfeinerung resultieren. Ebenso sind die bereits auf höherer Ebene modellierten Einstellungen als Verknüpfung an der Verfeinerung zu erfassen, falls dies gewünscht ist. Hierfür muss eine Zuordnung zur betreffenden Teilressource vorgenommen werden. Eine erneute Modellierung entfällt. Ähnliches gilt bei den Objekteigenschaften. Es gibt

somit verfeinerte Attribute, die zur nächst höheren Ebene verknüpft sind und andersherum, aber es kann ebenfalls Eigenschaften etc. geben, die nur in übergeordneter Ebene anfallen und keinen direkten Bezug in einer Verfeinerung aufweisen (z.B. Eig. z). Tabelle 2 fasst die Konzepte für die Datenerfassung bei verfeinerten Modellen zusammen.

Möglichkeit	**Probleme**	**Ansatz denkbar?**
1.) Erfassung auf einer Ebene	Ebenenauswahl Flexibilität des Systems beeinträchtigt mögliches Nichterfassen relevanter Werte, die auf anderer Ebene modelliert wurden	nein
2.) Erfassung auf verschiedenen Ebenen	u.U. keine ganzheitliche Betrachtung mehr möglich; Nachteile von 1.)	nein
3.) Erfassung auf verknüpften Ebenen	geeignetes Verknüpfungskonzept	ja

Tab. 2: Zusammenfassung der Möglichkeiten zur Datenerfassung bei verfeinerten Modellen

Sofern ein konsequentes Konzept für die interne Verknüpfung der Modellelemente der unterschiedlichen Ebenen vollständig und sinnvoll erarbeitet wurde, ist die Integration bestehender Datensätze genau so möglich wie eine Neuerfassung. Je nachdem auf welcher Ebene bereits erfasst wurde, kann es sein das einige Felder nach der Detaillierung leer sind. Tritt keine inhaltliche oder semantische Änderung auf, kann es wie bei einer regulären Modellmodifikation als neue Modellversion gehandhabt werden, die zu einer bestehenden Auswertungsbasis hinzugefügt wird (vorausgesetzt in der Modellierung wurden alle Referenzen gesetzt).

Überlegungen zu Anforderungen an eine technische Umsetzung

Für eine technische Umsetzung im Modellierungswerkzeug ergeben sich somit neue Anforderungen.

- Zur Wahrung der Übersichtlichkeit sollten Auswahlmöglichkeiten in der Anzeige eingeführt werden, die es erlauben Eigenschaften aus detaillierteren Prozessebenen aus- und einzublenden oder gar die gesamte Modellstruktur aus- bzw. einklappbar darzustellen.
- Bei der Erstellung einer Verfeinerung sollten bereits vorhandene Elemente zur Auswahl stehen, damit diese als Referenzen eingefügt werden können.
- Referenzierte Objekte und Eigenschaften sollten gekennzeichnet werden.

Außerdem wurden folgende Regeln (für Modellierung und Methode), aus den Überlegungen zur Integration des Verfeinerungskonzeptes in die derzeitige Methodenumgebung heraus, formuliert:

- Objekte am Anfang und am Ende der Verfeinerung sind Referenzen der zum verfeinerten Prozess gehörigen Objekte der höheren Ebene
- es komme keine Eingangs- oder Ausgangsobjekte hinzu, die in der darüber liegenden Ebene nicht vorkommen
- alle Elemente sind nur einmal in der Datenbank abgelegt und ggf. mit Referenzen verknüpft

3.4 Risiken beim Umgang mit Verfeinerungen

Die Verwendung von Verfeinerungen in der Modellierung bietet auch Fehlerpotenzial. Der Modellierer sollte daher sowohl Domänenerfahrung aufweisen, um die Prozessbeschreibung zu verstehen, als auch ein Mindestmaß an Modellverständnis mitbringen. So ist genau zu unterscheiden, ob es sich um eine Verfeinerung oder eine „Verlängerung“ der Prozesskette handelt. Nicht jede Ausweitung des Prozesses ist gleichgesetzt mit einer Verfeinerung, es kann sich ebenfalls um eine Erweiterung um zusätzliche (reale, vom ursprünglichem Prozessschritt unabhängige) Prozessschritte handeln. In diesem Fall greift das bisherige KM und es wird eine neue Versionsfamilie und damit eine neue Auswertbasis generiert.

Ein weiterer Nachteil dieses Ansatzes (der zugleich ebenfalls Fehlerpotenzial birgt) ist, das die Verwendung mehrerer Verfeinerungsebenen schnell zu Unübersichtlichkeit führen kann, wenn eine Vielzahl bereits existierende Elemente vorliegt. Doppelmodellierungen oder Falschreferenzierungen können die Folge sein. Spätestens an dieser Stelle ist es sinnvoll eine Konsistenzkontrolle zu integrieren, die vermeintliche Doppelmodellierungen aufspürt. Um derartig detail- und umfangreiche Modelle handhaben zu können, ist die bereits erwähnte Selektierung der Anzeige bzw. das Arbeiten mit verschiedenen Sichten essentiell.

Die Möglichkeit der Verwendung verschiedener Sichten und Anzeigedetaillierungen ist ebenso für die Erfassung und die Analyse relevant. Hierfür sollte die Auswahl nur bedingt durch den Bediener erfolgen. Vielmehr ist es bereits Aufgabe des Modellierers (oder eines Projektmanagers), festzulegen wer welche Einzelheiten sehen und bearbeiten darf. Dabei können unterschiedliche Konzepte zur Rechtevergabe diskutiert werden. Je nach Einzelfall kann es bspw. relevant sein Maschinendaten einsehen (nicht ändern) zu können oder eben aus Geheimhaltungsgründen maschineninterne Werte komplett aus Anzeige und Erfassung auszuschließen.

4 Zusammenfassung und Ausblick

Ausgehend von einer kurzen Beschreibung einer im produktionstechnischen Umfeld entstanden Methodik für die effektive Entwicklung und Verbesserung von Prozessen, wurde erläutert, welche Rolle die Modellierung innerhalb dieses Konzeptes spielt und

wie diese eingesetzt wird. Die Modellierung ist ein Kernbestandteil der Methodik und auf ihr aufbauend, sind die weiteren Schritte der Datenerhebung und Auswertung zur Generierung und Nutzbarmachung von Prozesswissen erstmöglich. Diese ganzheitliche Betrachtungsweise führt dazu, dass Modifikationen an Modellen mit besonderer Achtsamkeit durchgeführt werden müssen. Modellabhängige Komponenten (Daten, Analysen, etc.) sind direkt betroffen. Am Beispiel der Modellverfeinerung (deren Ursachen ebenfalls identifiziert wurden) wurden die einzelnen Probleme und Auswirkungen sowohl für die Modellierung, als auch für die abhängigen Komponenten diskutiert. Abschließend wurden erste Überlegungen über den Umgang mit diesen Komplikationen theoretisch und umsetzungstechnisch angeführt. Weiterführend wird dieses Verfeinerungskonzept präzisiert, so dass es softwareseitig (in einer, im Projektrahmen entstandenen und durch eine Ausgründung weiterentwickelten Anwendung [Sy15]) im Kontext der Gesamtmethodik integriert werden kann. Hierfür werden derzeit (zusätzlich zu den hier getätigten Überlegungen) Anpassungen und Umstrukturierungen in der Datenbank und im Konfigurationsmanagement der Methode erarbeitet sowie Berechtigungskonzepte überprüft, geändert und ergänzt. Die Formulierung formaler Constraints ist ebenfalls Forschungsbestandteil. Hierfür ist jedoch zu beachten, dass die Modelle durch Ingenieure erstellt werden, die dadurch bereits außerhalb ihres gewohnten Arbeitsumfeldes agieren. Sie sollten sich demnach nicht zusätzlich mit Constraints beschäftigen müssen. Eine automatisierte Umsetzung wird daher angestrebt.

Danksagung

Die hier vorgestellten Arbeiten wurden durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen des Sonderforschungsbereiches SFB 639 „Textilverstärkte Verbundkomponenten für funktionsintegrierende Mischbauweisen bei komplexen Leichtbauanwendungen“ gefördert.

Literaturverzeichnis

[Ba96] Baumgarten, B.: Petri-Netze – Grundlagen und Anwendungen, Spektrum Akademischer Verlag, 1996

[Du15] Duden online: www.duden.de (April 2015)

[DW97] Dangelmaier, W.; Warnecke, H.-J.: Fertigungslenkung. Planung und Steuerung des Ablaufs der diskreten Fertigung. Springer-Verlag, Berlin Heidelberg, 1997

[Fr99] Frank, U.: Zur Verwendung formaler Sprachen in der Wirtschaftsinformatik: Notwendiges Merkmal eines wissenschaftlichen Anspruchs oder Ausdruck eines übertriebenen Szientismus?, in (Becker, J. et al., Hrsg): Bestandsaufnahme und Perspektiven. Wiesbaden: Gabler 1999, S. 127-160

[Gr10] Großmann, K. et al.: Reproduzierbare Fertigung in innovativen Prozessketten. Konzeption eines Beschreibungs- und Analysetools (Teil 2). ZWF Zeitschrift für den wirtschaftlichen Fabrikbetrieb 105 (11), S.: 954-958, 2010

[Gr11] Großmann, K. et al.: Adopting potentials of biogenetic materials into composites by using the advantages of database-driven modelling. In Proc of International Conference on Innovative Technologies IN-TECH 2011

[GW10] Großmann, K.; Wiemer, H.: Reproduzierbare Fertigung in innovativen Prozessketten. Besonderheiten innovativer Prozessketten und methodische Ansätze für ihre Beschreibung, Analyse und Führung (Teil 1). ZWF Zeitschrift für den wirtschaftlichen Fabrikbetrieb 105 (10), S.: 855-859, 2010

[GW11] Großmann, K.; Wiemer, H.: Reproduzierbare Fertigung in innovativen Prozessketten. Anwendungsbeispiele und Erfahrungen (Teil 3). ZWF Zeitschrift für den wirtschaftlichen Fabrikbetrieb 106 (1-2), S.: 26-29, 2011

[JD02] Jacobs, H.-J.; Dürr, H.: Entwurf und Gestaltung von Fertigungsprozessen: Planung und Steuerung der spanenden Teilefertigung. Fachbuchverlag Leipzig, Carl Hanser Verlag München Wien. 2002

[OM13] OMG Unified Modeling Language (OMG UML) Version 2.5, September 2013; Download: http://www.omg.org/spec/UML/2.5

[Sc98] Schütte, R.: Grundsätze ordnungsmäßiger Referenzmodellierung: Konstruktion konfigurations- und anpassungsorientierter Modelle. Gabler, 1998

[Sc13] Scheer, A.-W.: ARIS – Vom Geschäftsprozess zum Anwendungssystem, Ausgabe 4, Springer Verlag, 2013

[St74] Stachowiak, H.: Allgemeine Modelltheorie. Springer, Wien, 1973

[Sy15] Symate GmbH: Software Detact; Webseite: http://www.symate.de/de/detact.html (Juni 2015)

[Th07] Thomas, O.: Version Management for Reference Models: Design and Implementation. In (Becker, J.; Delfmann, P., Hrsg.): Reference Modeling, 2007

[Wi13] Wiemer, H.: Technologiedatenmanagement als Werkzeug für Technologieentwicklung und Wissenstransfer. In Proc. Internationales ECEMP-Kolloquium: Ressourcenschonende Werkstoffe – Technologien – Prozesse, S.: 281-297. Dresden. Verlag wissenschaftliche Scripten. 2013

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Ein Modell zur Beschreibung des Interaktionsablaufs zwischen Mensch und Computer bei der Lösung von Planungsproblemen

Anna Prenzel[1]

Abstract: Planungsprobleme, z.B. aus den Gebieten der Produktions-, Flotten- oder Personalplanung können mit Hilfe von Software in vielen Fällen nahezu vollständig automatisch gelöst werden. Trotzdem wird in der Regel auch ein menschlicher Disponent benötigt, der sicherstellt, dass bei der Planung alle betriebs- und situationsspezifischen Anforderungen erfüllt werden. Die Lösung eines Planungsproblems wird damit zu einem integrierten Prozess, an dem Mensch und Computer beteiligt sind. In diesem Papier wird ein Interaktionsmodell entwickelt, welches unabhängig vom konkreten Anwendungsfall das dabei stattfindende Zusammenspiel zwischen den beiden Parteien beschreibt. Die Aktivität des Disponenten wird dabei als schrittweiser Vorgang der Variablenbelegung für das zugrunde liegende Constraint-Erfüllungsproblem (CSP) aufgefasst. Das Modell erlaubt es, Faktoren zu identifzieren, die den Arbeitsaufwand für den Disponenten stark erhöhen können. Dazu gehört u.a. die Behandlung von Sackgassen bei der Variablenbelegung des CSPs und die Berücksichtigung von Optimierungszielen. Es wird geschlussfolgert, dass der für den Disponenten entstehende Arbeitsaufwand durch eine geeignete Computerunterstützung reduziert werden sollte.

Keywords: Planungsprobleme, Lösungsverfahren, Expertenwissen, Mensch-Computer-Interaktion, Backtracking, CSP

1 Einführung

1.1 Anwendungsgebiet

In diesem Papier werden Feinplanungsprobleme betrachtet, mit denen die Zuordnung von Aufgaben zu Startzeitpunkten und Ressourcen geregelt wird. Beispiele sind die Zuordnung von Transportgütern zu Fahrern, Schichten zu Personal oder Fertigungsaufträgen zu Maschinen. Es wird angenommen, dass die Planung für einen begrenzten Planungszeitraum P mit $h \in \mathcal{N}^+$ diskreten Zeitpunkten $P = \{1,..,h\}$ vorgenommen wird, für den eine feststehende Nachfrage an Gütern oder Dienstleistungen vorliegt. Die Nachfrage bestimmt eine Menge von n Aufgaben $AS = \{a_1,..,a_n\}$, die im Rahmen der Planung bestimmten Ressourcen und Ausführungszeitpunkten zugeordnet werden müssen. Es seien m verschiedene Ressourcen $RS = \{r_1,..,r_m\}$ vorhanden. An der Ausführung einer Aufgabe ist jeweils eine Ressource oder eine Kombination von Ressourcen beteiligt. Es sei RCS die Menge möglicher Ressourcen und Ressourcenkombinationen $RCS \subseteq \mathcal{P}(RS)$.

[1] Hochschule Zittau/Görlitz, Fakultät Elektrotechnik und Informatik, Brückenstraße 1, 02826 Görlitz, aprenzel@hszg.de

Das Planungsproblem kann als Constraint-Erfüllungsproblem (CSP) modelliert werden [HW07, Dec03]. Ein CSP wird durch ein Tripel (V,D,C) beschrieben. Dabei ist V eine Menge von Entscheidungsvariablen, D die Menge der Domänen der Variablen und C die Menge aller Randbedingungen (Constraints) über diesen Variablen. Sei VS mit $|VS| = n$ die Menge aller Variablen $start_i$, die den Startzeitpunkt der Aufgabe a_i repräsentieren und VR mit $|VR| = n$ die Menge aller Variablen res_i, die die Ressourcenzuweisung von $a_i, i \in \{1..n\}$ festlegen. Dann gilt im Allgemeinen $(VS \cup VR) \subseteq V$. Es können weitere problemspezifische Variablen vorliegen.
Die Zuordnung von Variablen zu Domänen wird durch die Abbildung $dom : V \to D$ geregelt. Für alle $v \in VS$ gilt: $dom(v) \subseteq P$ und für alle $v \in VR$ gilt: $dom(v) \subseteq RCS$. Jeder Variable $v \in V$ dürfen nur Werte zugewiesen werden, die in der zugehörigen Domäne $dom(v)$ enthalten sind. Für die Wertzuweisung werden die Abbildungen $\alpha_i : v_i \to dom(v_i), i \in \{1..|V|\}$ definiert.
Jede Randbedingung $c \in C$ wird über einer Sequenz $Sub_c = \langle v_1,..,v_s \rangle, v_j \in V, j \in \{1..s\}, s \in \mathcal{N}^+$ definiert. Dabei entspricht c einer Relation, d.h. einer Teilmenge des Kreuzproduktes der beteiligten Variablen: $c \subseteq dom(v_1) \times ... \times dom(v_s)$. Typische Randbedingungen bei Feinplanungsproblemen sind z.B. Reihenfolgeabhängigkeiten zwischen Aufgaben, Zeitfenster für die Startzeitpunkte von Aufgaben oder Beschränkungen bei der Ressourcenzuweisung von Aufgaben. Eine Lösung für ein Planungsproblem liegt dann vor, wenn für alle $c \in C$ gilt: $Sub_c = \langle v_1,..,v_s \rangle \Rightarrow (\alpha(v_1),..,\alpha(v_n)) \in c$. In der Praxis wird meist keine beliebige Lösung gesucht, sondern eine Lösung mit einem optimalen Zielfunktionswert, der sich aus den Werten der Variablen berechnet.
Problemklassen, die dieser Modellstruktur gehorchen, sind z.B. die Ressourcenbeschränkte Projektplanung (RCPSP, [BK12]) , die Werkstattplanung (JSSP, [BK12]) oder die Flottenplanung (VRPTW, [KLMS05]).

1.2 Ein allgemeines Verfahren zur Lösung von Planungsproblemen

Ein weit verbreitetes Basisverfahren zur automatischen Lösung von Planungsproblemen als CSP ist eine Tiefensuche mit chronologischem Backtracking [HW07]. In Algorithmus 1 wird das zugrunde liegende Lösungsprinzip dargestellt.
In jedem Rekursionsdurchgang wird mit einer problemspezifischen Variablenauswahlheuristik $select\_var : \mathcal{P}(V) \to V$ eine noch nicht zugewiesene Variable ausgewählt. Anschließend wird mit Hilfe einer Wertauswahlheuristik $select\_val_s : v_s \to dom(v_s), s \in \{1..|V|\}$ ein Wert ausgewählt, sodass *assigned* eine partielle Lösung [Dec03] bildet. Wenn es nicht gelingt, in den nächsten Rekursionsschritten die restlichen Variablen in *unassigned* mit Werten zu belegen, wird für v_s eine neue Wertauswahl vorgenommen. Dies wird solange wiederholt, bis alle nachfolgenden Rekursionsschritte erfolgreich sind (und somit eine Lösung B existiert) oder bis die Domäne d_s leer ist. In diesem Fall wird v_s wieder als nicht zugewiesen betrachtet (Zeile 20) und der Wert der zuvor zugewiesenen Variable wird geändert (Backtracking). Wenn die Domäne d der im ersten Aufruf von *backtrackSolve* betrachteten Variable leer ist, existiert keine Lösung für das CSP.

Algorithmus 1 : Lösung eines CSPs mit Tiefensuche und Backtracking

input : Ein CSP (V,D,C)
output : Eine Lösung mit Wertzuweisungen für jede Variable, wenn eine Lösung existiert,
false, wenn keine Lösung existiert.

```
1  begin
2      unassigned ← V;
3      assigned ← ∅;
4      return backtrackSolve(assigned,unassigned);
5  end
6  backtrackSolve(assigned, unassigned) ≡
7  if unassigned ≠ ∅ then
8      v_s ← select_var(unassigned);
9      d_s ← dom(v_s);
10     while d_s ≠ ∅ do
11         value ← select_val_s(v_s);
12         d_s ← d_s \ {value};
13         v_select ← value;
14         assigned ← assigned ∪ {v_s};
15         unassigned ← unassigned \ {v_s};
16         B ← backtrackSolve(assigned, unassigned);
17         if B ≠ false then
18             return B;
19         else
20             assigned ← assigned \ {v_s};
21             unassigned ← unassigned ∪ {v_s};
22         end
23     end
24     returnfalse;
25 else
26     return assigned;
27 end
```

1.3 Die Beteiligung des Disponenten am Lösungsprozess

1.3.1 Gründe für die Beteiligung des Disponenten

Ein Planungsproblem ist strukturiert, wenn anhand des vorliegenden Modells (z.B. in Form eines CSPs) ohne Beteiligung des Disponenten automatisch ein Plan so erzeugt werden kann, dass alle betrieblichen Anforderungen erfüllt werden. In der Praxis sind die Probleme jedoch häufig semi-strukturiert, da einige Anforderungen nicht im Modell kodiert sind

[CvW12]. Dafür gibt es verschiedene Gründe:

1. Die Planungssoftware ist nicht individuell auf das Unternehmen oder die Organisation zugeschnitten, sondern wurde für einen breiten Anwenderkreis entwickelt (Standardanwendungen). Einige unternehmensspezifische Parameter, Variablen und Randbedingungen werden daher nicht bei der Planerstellung berücksichtigt.

Beispiel 1 (unternehmensspezifische Regeln) [dSvWJ11]:

- Fest angestellte Arbeiter sollten möglichst nicht in der gleichen Schicht wie befristet angestellte Mitarbeiter arbeiten.
- Fahrzeiten zwischen zwei Orten dürfen nicht mehr als 30 min. betragen.

2. Auch unternehmensspezifische Anwendungen bilden die Anforderungen häufig nicht vollständig ab, da es zum einen nicht immer möglich ist, während der Software-Herstellung alle Anforderungen zu ermitteln [Ceg08, CMWV99] und zum anderen nach der Einführung der Software neue Anforderungen entstehen können. Unspezifizierte Anforderungen umfassen häufig auch inoffizielle Planungsregeln: Der Disponent kennt aus Erfahrung die besten Ausführungszeitpunkte und Ressourcen für bestimmte Aufgaben.

Beispiel 2 (inoffizielle Regeln):

- Eine Aufgabe, deren Dauer mehr als 2 Stunden beträgt, sollte nicht zu Beginn des Arbeitstages eingeplant werden.
- Alle Aufgaben für Kunde X sollten von Arbeiter Y ausgeführt werden.

3. Das Planungsmodell wird bewusst gegenüber der Realität vereinfacht, um die Komplexität des Problems zu reduzieren und somit eine Lösungserstellung in akzeptabler Zeit zu ermöglichen. Es wird davon ausgegangen, dass der Disponent die zusätzlichen Anforderungen in einer automatisch generierten Ausgangslösung leicht manuell umsetzen kann [Ceg08].

4. Bestimmte Anforderungen können nur schwer kodiert werden. Nur der menschliche Disponent kann einschätzen, ob ein gegebener Plan zu einem bestimmten Zeitpunkt diese Anforderungen erfüllt oder nicht [dSvWJ11, CvW12].

Beispiel 3 (schwer kodierbare Anforderungen):

- Planänderungen sollten nachvollziehbar sein.
- Pläne sollten für die Mitarbeiter akzeptabel sein.
- Die Aufgaben sollten möglichst abwechslungsreich verteilt werden. [dSvWJ11].

6. Unabhängig von den restlichen Punkten wird das menschliche Urteilsvermögen des Disponenten benötigt, um Unsicherheiten in den Daten zu identifizieren und entsprechende Maßnahmen einleiten zu können [CvW12].

Beispiel 4 (vorausschauende Planung): Der Disponent einer Produktionsfirma erwartet eine Materiallieferung, die für nächste Woche angekündigt ist. Bei einer pünktlichen Lieferung könnte die Verarbeitung in der darauffolgenden Woche erfolgen. Da sich der Lieferant erfahrungsgemäß verspätet, plant sie der Disponent jedoch erst in der übernächsten Woche ein.

1.3.2 Die Umsetzung manueller Planungsentscheidungen

Es lassen sich zwei Arten von Planungsentscheidungen identifizieren, die vom Disponenten zur Umsetzung nicht kodierter Anforderungen getroffen werden können:

- Der Disponent beschränkt die Startzeitpunkte bestimmter Aufgaben jeweils auf bestimmte Bereiche im Planungshorizont.
- Der Disponent beschränkt die Ressourcenzuweisung bestimmter Aufgaben jeweils auf bestimmte Gruppen von Ressourcen.

Diese Planungsentscheidungen können auf zwei Wegen in den Planungsprozess eingebracht werden:

1. *Überführung in kodierte Anforderungen:* Dabei wird das Planungsmodell um unäre Randbedingungen erweitert, die den Wertebereich ausgewählter Variablen auf eine Teilmenge des ursprünglichen Wertebereiches (evtl. auf einen einzigen Wert) einschränken. Das erweiterte Planungsproblem kann anschließend ggf. automatisch gelöst werden.
2. *Manuelle Modifikation einer automatisch erzeugten Lösung:* Dabei werden die Werte von einzelnen (Startzeit- und Ressourcen-)Variablen in einer Lösung nachträglich geändert.

In vielen Fällen kann Expertenwissen nicht sofort in passende Planungsentscheidungen übersetzt werden [BPYS07]. Häufig ist nicht von vornherein bekannt, mit welchen Entscheidungen eine Anforderung am besten umgesetzt werden kann. In diesem Fall muss der

Disponent alternative Randbedingungen bzw. Wertzuweisungen aufstellen und bewerten. Dieser Vorgang lässt sich als 3-stufiger Entscheidungsprozess mit den Stufen *Intelligenz*, *Entwurf* und *Auswahl* modellieren [TSD11]:

Intelligenz: Die Intelligenzphase ist der Einstieg in den Entscheidungsprozess. In einer bestimmten Situation stellt der Disponent die Notwendigkeit fest, die Eigenschaften des Plans zu beeinflussen. Er sammelt Wissen, welches zunächst in Form von nicht kodierten Anforderungen vorliegt.

Entwurf: In dieser Phase werden Planungsentscheidungen gesucht, mit denen sich die Anforderungen beschreiben lassen. Dabei werden Alternativen entworfen, bewertet und miteinander verglichen. Folgende Kriterien fließen in die Bewertung ein:

- Wie gut werden die Anforderungen durch diese Entscheidungen repräsentiert?
- Existiert eine Lösung mit einer ausreichend guten Qualität, wenn die Entscheidungen im Modell bzw. in der Lösung umgesetzt werden?
- Wie gut ist der Kompromiss, der mit diesen Entscheidungen zwischen den einzelnen (ggf. widersprüchlichen) Anforderungen erreicht werden kann?

Auswahl: In der letzten Phase wird eine Alternative ausgewählt, d.h. es werden die Entscheidungen bestimmt, mit denen die Anforderungen am besten abgebildet werden können.

Der Entwurfsprozess kann übersprungen werden, wenn sich die Anforderungen direkt in Planungsentscheidungen überführen lassen.

2 Beschreibung des interaktiven Lösungsprozesses für Planungsprobleme

Der Disponent muss mit dem Planungssystem interagieren, um einzelne Planungsentscheidungen in den Planungsprozess einbringen zu können. Dazu gehört z.B. das Eintragen von Randbedingungen an der Benutzerschnittstelle oder das Ändern der Position einer Aufgabe im Plan (häufig können Aufgaben in einer interaktiven grafischen Visualisierung des Plans mittels Drag-and-Drop verschoben werden). Das System verarbeitet diese Aktionen und aktualisert das Planungsergebnis entsprechend, woraus der Disponent ggf. wiederum Schlussfolgerungen für neue oder geänderte Planungentscheidungen zieht. Es wird deutlich, dass im Rahmen des 3-stufigen Entscheidungsprozesses eine intensive Interaktion zwischen Mensch und Computer stattfinden kann. In diesem Abschnitt wird ein Modell zur Beschreibung der möglichen Interaktionsabläufe entwickelt.

2.1 Annahmen an das Planungssystem

Im nachfolgenden Szenario wird angenommen, dass das verwendete Planungssystem folgende Funktionalität bereitstellt:

- eine automatische Planerstellung, bei der (falls möglich) eine Lösung unter Einbeziehung aller Aufgaben oder wahlweise eine partielle Lösung unter Einbeziehung einer Teilmenge von Aufgaben erzeugt wird,
- die Möglichkeit zur Festlegung unärer Randbedingungen durch den Disponenten,
- die Möglichkeit der manuellen Wertzuweisung für beliebige Variablen durch den Disponenten (z.B. durch Verschieben einer Aufgabe in der grafischen Planansicht),
- die Möglichkeit, Wertzuweisungen für bestimmte Variablen offen zu lassen oder zu löschen, sodass eine partielle Instanziierung der Variablen vorliegt,
- die Benachrichtigung des Disponenten, wenn Randbedingungen verletzt werden,
- die Berechnung der Qualität des Plans in Form von Leistungskennzahlen.

2.2 Der übergeordnete Entscheidungsprozess

Algorithmus 2 modelliert den Ablauf des interaktiven Lösungsprozesses aus einer abstrakten Sichtweise. Die Zeilen 2 bis 14 umfassen den übergeordneten Entscheidungsprozess, in dem mehrere Alternativlösungen entwickelt werden, bis der Disponent den Vorgang abbricht. Jede Alternativlösung wird in den Zeilen 3 bis 13 erzeugt, was als eingebetteter Entscheidungsprozess aufgefasst werden kann:

Zeilen 3-7: Zunächst entscheidet der Disponent, ob vom Computer eine vollständige Ausgangslösung erzeugt werden soll ($userInitialize = true$) oder nicht ($userInitialize = false$). Durch die Prozedur *computerSolve* wird ein beliebiger Lösungs- oder Optimierungsalgorithmus repräsentiert.

Zeilen 8-13: Wenn gilt $userInitialize = true$ und eine Lösung existiert, dann sind anschließend alle Variablen zugewiesen (und liegen in Liste *assigned* vor). Ansonsten sind noch keine Variablen zugewiesen (*assigned* ist leer). Der Disponent hat nun die Möglichkeit, neue Zuweisungen vorzunehmen oder bestehende Zuweisungen zu ändern oder zu lösen. Die Prozedur *userSolve* repräsentiert dabei den Entwurf einer alternativen (ggf. partiellen) Lösung, in der bestimmte Anforderungen umgesetzt werden sollen. Wenn die erzeugte Lösung noch unvollständig ist, kann sie der Disponent automatisch vervollständigen lassen (Zeile 13). Die in *assigned* bereits vorliegenden automatischen oder manuellen Wertzuweisungen werden dabei als Randbedingungen berücksichtigt.

Jede Alternative kann iterativ angepasst werden, bis der Disponent den Vorgang abbricht ($userCancelled = true$) und ggf. ab Zeile 2 die Umsetzung einer neuen Alternative beginnt. Eine weitere Iteration ist z.B. erforderlich, wenn durch *computerSolve* festgestellt wird, dass für die in *assigned* vorgegebenen Zuweisungen keine Lösung existiert, sodass diese nochmals angepasst werden müssen. Auch wenn der Disponent aus der vervollständigten Lösung neue Präferenzen für die Lösungseigenschaften ableitet, ist eine weitere Iteration erforderlich.
Die automatische Planung kann vom Nutzer konfiguriert werden (Liste *settings* als Para-

meter für *computerSolve*). Dies ermöglicht z.B. die Wahl eines Optimierungsverfahrens oder die Gewichtung von Zielfunktionen.

Algorithmus 2 : Interaktive Lösung eines CSPs für die Planung (Teil 1)

input : Ein CSP (V,D,C)
output : Eine Lösung mit Wertzuweisungen für jede Variable, wenn eine Lösung existiert,
false, wenn keine Lösung existiert.

```
1  begin
2      while !userCancelled do
3          unassigned ← V;
4          assigned ← ∅;
5          /*vgl. Algorithmus 3:*/
6          if userInitialize then
7              computerSolve(assigned, unassigned, settings);
8          end
9          while !userCancelled do
10             /*Zuweisungen manuell vervollständigen, ändern oder lösen:*/
11             userSolve(assigned, unassigned);
12             /*übrige Variablen in Liste unassigned zuweisen lassen:*/
13             computerSolve(assigned, unassigned, settings);
14         end
15     end
16 end
```

2.3 Der Prozess zum Entwurf einer Alternativlösung

In diesem Abschnitt steht der Entwurf einer einzelnen Alternativlösung im Vordergrund. Es wird davon ausgegangen, dass der Disponent eine Kombination von Planungsentscheidungen sucht, mit der eine oder mehrere Anforderungen beschrieben werden können. Der Prozess zur sukzessiven Eingabe der Entscheidungen an der Schnittstelle des Planungssystems besitzt Gemeinsamkeiten mit der Tiefensuche. Im Gegensatz zu Algorithmus 1 wird dabei sowohl die Variablen- als auch die Wertauswahl durch den Disponenten vorgenommen. Der vollständige Ablauf der Variablenbelegung wird in Algorithmus 3 nach dem Vorbild der Tiefensuche als rekursive Funktion modelliert[1]: In jedem Rekursionsdurchgang wird zunächst eine beliebige bereits zugewiesene oder noch nicht zugewiesene Variable ausgewählt (Zeile 2). Eine zugewiesene Variable kann vom Disponenten sofort in die Liste noch nicht zugewiesener Variablen verschoben werden (Zeilen 3-5). Dies ist erforderlich, wenn die Wertzuweisung für diese Variable nach der Entwurfsphase dem

[1] **Konventionen für die folgenden Interaktions-Algorithmen:** Funktionen mit dem Präfix *user* repräsentieren Entscheidungen oder (z.T. rein kognitive) Aktivitäten des Disponenten. Der Präfix *computer* kennzeichnet eine unterstützende Funktion, die vom Disponenten an der Benutzerschnittstelle aufgerufen werden kann. Im Kopf einer Bedingung (if..then) werden z.T. Variablen mit dem Präfix *user* eingesetzt, deren Wert für eine Ja/Nein-Entscheidung des Disponenten steht.

Computer überlassen werden soll (Algorithmus 2, Zeile 13) oder wenn der Disponent diese Variable zunächst ignorieren möchte, damit er andere Variablenzuweisungen leichter (d.h. ohne Verletzung von Randbedingungen) vornehmen kann. Wenn die Variable nicht ignoriert wird, wird für sie in den Zeilen 6 bis 43 ein (neuer) Wert gesucht. Dieser Vorgang gestaltet sich wie folgt:

Zeilen 10-12: Der Disponent wählt eine Reihe von alternativen Werten oder Teilbereichen des Wertebereiches aus, die aus seiner Sicht für eine Zuweisung in Frage kommen.

Zeilen 14-20: Jeder Wert wird daraufhin überprüft, ob er gültig und zufriedenstellend ist. Ein Wert ist gültig, wenn seine Zuweisung keine Randbedingungen mit bereits zugewiesenen Variablen verletzt. Er ist zufriedenstellend, wenn die um diese Wertzuweisung ergänzte (ggf. partielle) Lösung die Anforderungen des Disponenten ausreichend erfüllt.

Zeilen 21-32: Wenn kein passender Wert existiert, kann der Disponent die Zuweisung dieser Variable abbrechen (Zeilen 22-29) und den Lösungsprozess mit einer anderen Variable fortzusetzen (Zeile 41) bzw. den Entwurf dieser Alternative komplett abbrechen (Zeile 39). Eine Variable, die bereits zugewiesen war, erhält dabei ihren alten Wert zurück (Zeile 24). Alternativ kann der Disponent einen Backtracking-Prozess einleiten, indem er einige der vorhandenen Wertzuweisungen ändert oder löst (dazu wird in Zeile 30 die Funktion *userSolve* rekursiv aufgerufen, wobei nur die bereits zugewiesenen Variablen als Parameter übergeben werden). Anschließend kann ab Zeile 13 erneut eine Überprüfung der ausgewählten Werte stattfinden.
Im Gegensatz zu Algorithmus 1 findet kein systematischer Backtracking-Vorgang statt, bei dem jeweils die letzte Variablenzuweisung geändert wird. Die menschliche Kapazität des menschlichen Kurzzeitgedächtnisses ist gegenüber dem Computer stark begrenzt [Mac08], sodass nicht davon ausgegangen werden kann, dass der Disponent die Reihenfolge der betrachteten Variablen beibehält.

Zeilen 32-36: Wenn mindestens ein Kandidat existiert, kann der Disponent einen Wert aus der Liste kandidierender Werte auswählen und der betrachteten Variable zuweisen.

Die Zuweisung zu jeder ausgewählten Variable kann als eingebetteter Entscheidungsprozess mit Entwurfs- und Auswahlphasen betrachtet werden, da sich der Disponent in der Regel zwischen alternativen Wertzuweisungen entscheiden muss.

Algorithmus 3 : Interaktive Lösung eines CSPs für die Planung (Teil 2)

```
1  userSolve(assigned, unassigned) ≡
2  /*Variablenauswahl aus assigned oder unassigned:*/
   V_select ← user_select_var(assigned, unassigned);
3  if userIgnores && V_select ∈ assigned then
4      assigned ← assigned \ {V_select};
5      unassigned ← unassigned ∪ {V_select};
6  else
7      if V_select ∈ assigned then
8          old_value ← α(V_select);
9      end
10     unassigned ← unassigned \ {V_select};
11     sub_domain ← user_select_values(D_select);
12     candidates ← ∅;
13     while candidates = ∅ do
14         for each value ∈ sub_domain do
15             V_select ← value;
16             if computer_is_feasible?(V_select) && user_is_satisfying?(V_select) then
17                 candidates ← candidates ∪ value;
18             end
19             V_select ← null;
20         end
21         if candidates = ∅ then
22             if userCancelledWhile then
23                 if V_select ∈ assigned then
24                     V_select ← old_value;
25                 else
26                     unassigned ← unassigned ∪ {V_select};
27                 end
28                 /*Sprung zu Zeile 38:*/ break;
29             else
30                 userSolve(assigned);
31             end
32         else
33             value ← user_select_val(candidates);
34             V_select ← value;
35             assigned ← assigned ∪ {V_select};
36         end
37     end
38     if userCancelled then
39         return;
40     else
41         userSolve(assigned, unassigned);
42     end
43 end
```

3 Bewertung verschiedener Interaktionsszenarien hinsichtlich des zu erwartenden Arbeitsaufwandes

Aus dem in Algorithmus 3 beschriebenen Interaktionsmodell können verschiedene konkrete Interaktionsszenarien abgeleitet werden. Typische Szenarien sind z.B.:

Manuelle Planung: Der Disponent nimmt alle Variablenzuweisungen selbst vor (in Algorithmus 2 gilt $userInitialize = false$).

Automatische Planung mit nachträglicher manueller Modifikation: Es wird eine Ausgangslösung erzeugt ($userInitialize = true$). Anschließend werden einige Zuweisungen manuell geändert.

Automatische Planung unter Berücksichtigung zusätzlicher Randbedingungen: Es wird keine Ausgangslösung erzeugt ($userInitialize = false$). Der Disponent nimmt einige manuelle Wertzuweisungen vor (dies entspricht der Festlegung von Randbedingungen) und lässt die Lösung anschließend automatisch vervollständigen.

Die Szenarien können miteinander kombiniert werden. Im Folgenden wird jedes Szenario auf Faktoren untersucht, die den Zeitaufwand und die Schwierigkeit der Arbeit des Disponenten erhöhen oder seine Motivation zur Beteiligung am Lösungsprozess beeinträchtigen können. Es wird vorausgesetzt, dass die in Abschnitt 2.1 beschriebenen Annahmen an das Planungssystem gelten. In Abschnitt 3.4 werden daraus Schlussfolgerungen für die Gestaltung der Mensch-Computer-Schnittstelle gezogen.

3.1 Manuelle Planung

Eine manuelle Planung ist für den Disponenten in der Regel mit einem hohen Arbeitsaufwand verbunden. Dies gilt besonders dann, wenn während der Zuweisung häufig Sackgassen auftreten. Eine Sackgasse liegt vor, wenn der Disponent die Spezifikation einer Alternativlösung nicht mit weiteren Planungsentscheidungen fortsetzen kann, ohne Randbedingungen zu verletzen ($is\_feasible$ ist stets $false$ in Algorithmus 3, Zeile 16). Der Disponent muss daher die bisher erstellte partielle Lösung abwandeln (Zeile 30).
Der Disponent hat in der Regel keine Möglichkeit, das Auftreten von Sackgassen durch eine vorausschauende Planung zu verhindern. Wenn er eine Planungsentscheidung trifft, kann er die Auswirkungen auf die restlichen Variablen nicht sofort feststellen, denn bei einer großen Menge an Variablen und Randbedingungen ist es ihm nahezu unmöglich, im Kopf eine vorausschauende Propagierung durchzuführen. Der Grund ist zum einen die begrenzte Kurzzeitgedächtniskapazität des Menschen [Mac08], die von der Größe praktischer Planungsprobleme in der Regel weit überschritten wird. Zum anderen sind die zugrunde liegenden Randbedingungen in der Regel unbekannt, sodass sie nicht in die Propagierung einbezogen werden können.
Ob der Wertebereich einer Variablen nach der Propagierung seiner Entscheidung keine Werte mehr enthält, kann der Disponent also erst dann feststellen, wenn er die betroffene

Variable betrachtet. Diese steht u.U. in der vom Disponenten gewählten Zuweisungsreihenfolge sehr weit hinten. Es finden also u.U. zahlreiche Belegungen nicht betroffener Variablen statt, bevor ein Konflikt bemerkt wird. An dieser Stelle ist die Konfliktursache jedoch meist nicht mehr nachvollziehbar, d.h. der Disponent kann die Variable, für die ein Konflikt vorliegt, nicht mehr der Variable zuordnen, die den Konflikt verursacht hat. Letztere muss, wenn der Wertebereich leer ist oder nicht mehr den gewünschten Wert enthält, in einem zeitaufwändigen Backtracking-Prozess herausgefunden werden.
In Abhängigkeit vom Modell und von der Anzahl der Variablen kann es für den Disponenten zudem schwierig bzw. unmöglich sein, neben der Erfüllung der Randbedingungen auch eine Optimierung der Zielfunktionen anzustreben.

3.2 Automatische Planung mit nachträglicher manueller Modifikation

Verschlechterung der Qualität
Nachträgliche lokale Modifikationen können die Qualität des Plans bezüglich ein oder mehrerer Kennzahlen verschlechtern. Der Disponent arbeitet nach dem Prinzip einer Verbesserungsheuristik: Jede Planänderung, bei der die Randbedingungen berücksichtigt werden, führt zu einer Nachbarschaftslösung mit einer besseren oder schlechteren Qualität. Auch wenn der Disponent neben der Umsetzung von Anforderungen auch auf die Optimierung der Zielfunktionen achtet, besteht die Gefahr, ein lokales Optimum zu erreichen. Häufig befindet sich im Lösungsraum des Planungsproblems ein Plan, der den Vorstellungen des Disponenten ebenfalls entspricht und gleichzeitig eine bessere Qualität besitzt. Es ist jedoch schwierig, diesen Plan manuell zu finden. In der Regel ist dazu eine größere Umstellung notwendig, die mit einem hohen Arbeitsaufwand verbunden ist.

Selbstgefälligkeit
Dieser Effekt bezeichnet das übermäßige Vertrauen eines Nutzers in ein automatisiertes System und die damit verbundene Verschlechterung des Expertenwissens. Er kann dazu führen, dass nach einer automatischen Planerstellung keine manuellen Modifikationen am Plan vorgenommen werden, obwohl noch nicht erfüllte Präferenzen vorhanden sind. Zum einen weiß der Disponent, dass es ihm (im Gegensatz zum Computer) schwerfällt, gute Qualitätskennzahlen des Gesamtplans zu erreichen. Zum anderen sind ihm u.U. nicht alle Randbedingungen und Regeln bekannt, die in das Planungssystem eingepflegt wurden. Es entsteht eine Unsicherheit darüber, ob wichtige Faktoren bei der geplanten Modifikation nicht berücksichtigt wurden. Die Befürchtungen, die Qualität zu verschlechtern oder Fehlentscheidungen zu treffen, lassen es letztendlich am sichersten erscheinen, die Planung vollständig dem Computer zu überlassen.

Aufwand der manuellen Planung → Abschnitt 3.1

3.3 Automatische Planung unter Berücksichtigung zusätzlicher Randbedingungen

Planung durch Versuch und Irrtum
Dem Disponenten wird keine Unterstützung dabei geboten, die Randbedingungen nach

dem Gesichtspunkt der Lösbarkeit auszuwählen. Die automatische Planung muss u.U. mehrfach mit angepassten Randbedingungen aufgerufen werden, bevor das Problem lösbar ist (mehrere Iterationen in Algorithmus 2, Zeilen 9-14). Der Arbeitsaufwand der Planung wird dadurch erhöht.

Mangelnde Kontrolle über das Ausmaß der Umplanung
Die nachträgliche Modifikation eines Plans kann die Anpassung von bestehenden Zuweisungen erfordern (Backtracking). Wenn der Disponent die Anpassung nicht manuell vornehmen möchte, kann er einige der bestehenden Zuweisungen lösen (Algorithmus 3, Zeilen 4-5) und den Plan nach seiner Modifikation automatisch vervollständigen lassen (Algorithmus 2, Zeile 13). Häufig ist es bei einer nachträglichen Änderung jedoch erwünscht, dass die bestehenden Planungsentscheidungen, abgesehen von den Modifikationen des Disponenten, weitestgehend unverändert beibehalten werden. Bei einem Backtracking-Vorgang strebt der Disponent daher die geringstmögliche Anpassung des Plans an, die zur Auflösung des Konfliktes erforderlich ist. Eine automatische Neuplanung passt den Plan jedoch u.U. stärker an als erwünscht, sodass sich weitreichende und schwer nachvollziehbare Planänderungen ergeben.

3.4 Rückschlüsse für die Gestaltung der Mensch-Computer-Schnittstelle

Aus der Untersuchung der Algorithmen 2 und 3 können Anforderungen für die Gestaltung der Mensch-Computer-Schnittstelle von Planungssystemen abgeleitet werden:

1. Das Planungssystem sollte den Disponenten während der Planung bei der Auswahl von Werten aus dem Wertebereich einer Entscheidungsvariable unterstützen. Werte, die zur Verletzung einer Randbedingung oder in zukünftigen Zuweisungsschritten in eine Sackgasse führen, sollten entsprechend gekennzeichnet werden.
2. Der Disponent sollte einen gewünschten Wert auch dann auswählen können, wenn er einen Konflikt verursacht, d.h. wenn er im Zusammenhang mit den bisherigen Zuweisungsschritten zur Verletzung einer Randbedingung oder in zukünftigen Zuweisungsschritten in eine Sackgasse führt. Der Backtracking-Prozess zur Beseitigung des Konflikts sollte vom Planungssystem übernommen bzw. unterstützt werden. Dabei sollte eine möglichst geringe Abweichung von den Entscheidungen des Disponenten angestrebt werden.
3. Das Planungssystem sollte die Möglichkeit bereitstellen, Planungsentscheidungen automatisch innerhalb von vom Disponenten vorgegebenen Grenzen so anzupassen, dass Qualität des Plans optimiert wird.

Die Umsetzung dieser Anforderungen hilft dabei, den Arbeitsaufwand der manuellen Planung zu reduzieren und die Qualität der interaktiv erstellten Pläne zu verbessern. Sie trägt damit zu einer erfolgreicheren und zufriedenstellenderen Beteiligung des Disponentem am Lösungsprozess bei, wodurch auch der Effekt der Selbstgefälligkeit reduziert werden kann.

4 Zusammenfassung

In diesem Papier wurde die Beteiligung des Disponenten am Prozess zur Lösung von Planungsproblemen untersucht. Es wurde ein mehrstufiger Entscheidungsvorgang identifiziert, in dem der Disponent Zuweisungen für Startzeiten und Ressourcen einzelner Aufgaben sucht, um bestimmte Anforderungen zu erfüllen, die nicht im Planungsmodell abgebildet sind. Der Vorgang wurde nach dem Vorbild einer Tiefensuche mit Backtracking zur Lösung eines CSPs modelliert. Aus dem Modell wurden verschiedene Interaktionsszenarien abgeleitet und nach dem vom Disponenten benötigten Arbeitsaufwand bewertet. Es wurden Aufwandsfaktoren und Interaktionsdefizite wie z.B. die Notwendigkeit für Backtracking-Vorgänge, die potenzielle Verschlechterung der Qualität und der Effekt der Selbstgefälligkeit identifiziert. Bei der Gestaltung von Planungssystemen müssen Maßnahmen getroffen werden, um diese Defizite so weit wie möglich zu kompensieren.

4.1 Verwandte Arbeiten

Viele Studien stimmen darin überein, dass die Beteiligung des Disponenten an der Planerstellung in der Praxis unerlässlich ist. Es wurden verschiedene Konzepte der Funktionsaufteilung zwischen Mensch und Computer entwickelt: Nach dem Konzept der *gemischten Initiative* müssen beide Parteien als gleichberechtigte Agenten betrachtet werden, die gemeinsam ein bestimmtes Planungsziel erreichen wollen und dazu in jeder Stufe des Lösungsprozesses ihr jeweiliges Wissen auf eigene Initiative einbringen [BBIM96]. Unter dem Begriff der *interaktiven Optimierung* wurden Varianten der Beteiligung des Disponenten konkretisiert. Dazu gehört:

- Die Einbeziehung des Disponenten in die automatische Lösungserstellung, z.B. um in einem lokalen Optimierungsverfahren die Suche nach der optimalen Lösung in eine bestimmte Richtung zu lenken (vgl. [KLMM10]).
- Die Beteiligung des Disponenten an der Modell- und Wissensspezifikation. Verschiedene Möglichkeiten zur Modelländerung, darunter auch die in Abschnitt 2.1 beschriebenen Interaktionsmöglichkeiten, werden von [dNE05] zusammengefasst. Ein Beispiel ist das in [SHC05] vorgestellte System COMIREM (Interaktive Ressourceneinsatzplanung).
- Die teilweise Übernahme der Lösungserstellung durch den Anwender.

Eine Analyse der Nutzeraktionen im Kontext eines vollständigen Interaktionsvorgangs von der Feststellung des Handlungsbedarfs bis zur Fertigstellung des gewünschten Plans findet dabei nicht statt. Mögliche Interaktionsprobleme, die sich aus der Größe von Such- und Lösungsraum praktischer kombinatorischer Planungsprobleme ergeben, werden dadurch nicht identifiziert. In diesem Papier wird eine mögliche Vorgehensweise zur Durchführung einer solchen Analyse gezeigt.

Literatur

[BBIM96] Mark H. Burstein, Bolt Beranek, Newman Inc und Drew V. Mcdermott. Issues in the development of human-computer mixed-initiative planning. In *Cognitive Technology*, Seiten 285–303. Elsevier, 1996.

[BK12] Peter Brucker und Sigrid Knust. Scheduling Models. In *Complex Scheduling*, GOR-Publications, Seiten 1–28. Springer Berlin Heidelberg, 2012.

[BPYS07] Pauline Berry, Bart Peintner und Neil Yorke-Smith. Bringing the User Back into Scheduling: Two Case Studies of Interaction with Intelligent Scheduling Assistants. In *Interaction Challenges for Intelligent Assistants*, Seiten 10–11, 2007.

[Ceg08] Julien Cegarra. A cognitive typology of scheduling situations: A contribution to laboratory and field studies. *Theoretical Issues in Ergonomics Science*, 9(3):201–222, 2008.

[CMWV99] S. Crawford, B. L. MacCarthy, J. R. Wilson und C. Vernon. Investigating the Work of Industrial Schedulers through Field Study. *Cognition, Technology and Work*, 1:63–77, 1999.

[CvW12] Julien Cegarra und Wout van Wezel. Revisiting Decision Support Systems for Cognitive Readiness: A Contribution to Unstructured and Complex Scheduling Situations. *Journal of Cognitive Engineering and Decision Making*, 6(3):299–324, 2012.

[Dec03] R. Dechter. *Constraint Processing*. The Morgan Kaufmann Series in Artificial Intelligence. Elsevier Science, 2003.

[dNE05] Hugo A.D. do Nascimento und Peter Eades. User hints: a framework for interactive optimization. *Future Generation Computer Systems*, 21(7):1177–1191, 2005.

[dSvWJ11] Cees de Snoo, Wout van Wezel und René J. Jorna. An empirical investigation of scheduling performance criteria. *Journal of Operations Management*, 29(3):181–193, 2011.

[HW07] P. Hofstedt und A. Wolf. *Einführung in die Constraint-Programmierung: Grundlagen, Methoden, Sprachen, Anwendungen*. Springer, 2007.

[KLMM10] Gunnar W. Klau, Neal Lesh, Joe Marks und Michael Mitzenmacher. Human-guided search. *Journal of Heuristics*, 16(3):289–310, 6 2010.

[KLMS05] Brian Kallehauge, Jesper Larsen, OliB.G. Madsen und MariusM. Solomon. Vehicle Routing Problem with Time Windows. In Guy Desaulniers, Jacques Desrosiers und MariusM. Solomon, Hrsg., *Column Generation*, Seiten 67–98. Springer US, 2005.

[Mac08] M. MacDonald. *Your Brain: The Missing Manual*. Missing manual. O'Reilly Media, 2008.

[SHC05] Stephen F. Smith, David W. Hildum und David R. Crimm. Comirem: An Intelligent Form for Resource Management. *IEEE Intelligent Systems*, 20, 2005.

[TSD11] Efraim Turban, Ramesh Sharda und Dursun Delen. *Decision Support and Business Intelligence Systems*. Pearson, 2011.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Produktionsplanung mit einer kombinierten Lösung aus mathematischer Optimierung und anderen Business Analytics Werkzeugen

Hans Schlenker[1]

Abstract: Produktionsplanung ist eine bedeutende und komplexe Planungsaufgabe die in sehr vielen Unternehmen durchgeführt wird. Häufig wird diese Planung auch heute noch überwiegend manuell durchgeführt. Im vorliegenden Papier wird ein neuer Software-Ansatz skizziert, der aktuelle Business Analytics Ansätze kombiniert: statistische Vorhersage, mathematische Optimierung und Business Intelligence. Dabei liegt der Fokus auf den verwendeten Werkzeugen und wie sie zusammenarbeiten. Die Lösung wird abschließend anhand einer konkreten Fallstudie bewertet.

Keywords: Business Analytics, mathematische Optimierung, statistische Methoden, Vorhersage, IBM, ILOG, CPLEX, SPSS, Cognos TM1.

1 Einleitung

Mathematische Optimierung, statistische Vorhersage und *Business Intelligence* gelten als drei Grundpfeiler von *Business Analytics*. Business Intelligence Lösungen erlauben Entscheidern, ihre Daten intelligent zu bearbeiten, zu visualisieren und zu analysieren. Statistische Vorhersage liefert automatische Prognose auf der Basis historischer Daten, mittels Verfahren aus der mathematischen Statistik. Und mathematische Optimierung schließlich verwendet insbesondere die prognostizierten Zukunftsdaten um konkrete Handlungsempfehlungen für die Zukunft optimal zu berechnen.

Im Folgenden wird gezeigt wie diese Verfahren in aktuellen Software-Werkzeugen eingesetzt werden und wie sie technisch zusammenspielen. Schließlich wird ein konkretes Anwendungsszenario skizziert und aufgezeigt, welchen fachlichen Mehrwert die technische Lösung dort bringt.

2 Daten-Analyse und Statistische Vorhersage

Der erste Schritt der Produktionsplanung ist: Datenanalyse und Vorhersage. Mit dem Werkzeug *IBM SPSS Modeler* können Daten aus unterschiedlichen Quellen gelesen, analysiert und auch verändert werden, beispielsweise zur Behebung von Datenfehlern. Der Daten-Analyst definiert im SPSS Modeler einen sog. Daten-Stream, der die

[1] IBM Deutschland GmbH, Hollerithstr. 1, 81829 München, hans.schlenker@de.ibm.com

Eingabedaten verarbeitet. Abb. 1 zeigt einen solchen Stream im Modeler.

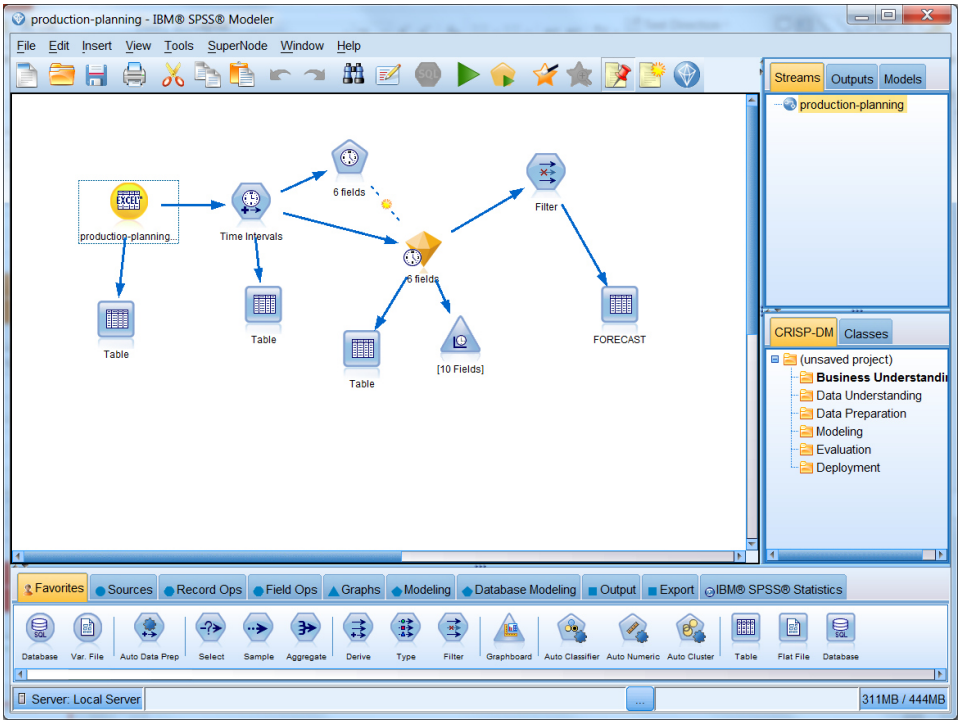


Abb. 1: Ein Daten-Stream in *IBM SPSS Modeler* zur Vorhersage von Verkaufszahlen.

Dieser Stream liest historische Verkaufszahlen aus einer Excel-Datei (dieser Startpunkt ist durch den gelb dargestellten Kreis gegeben), erzeugt damit eine neue Zeitreihe, erweitert diese Zeitreihe um weitere Perioden in die Zukunft, wendet dann Vorhersagemodelle an, um für die zukünftigen Zeitperioden möglichst wahrscheinliche Verkaufszahlen zu ermitteln, filtert diese Zahlen und schreibt sie schließlich in eine Tabelle FORECAST. Die Vorhersagemodelle werden vom Werkzeug SPSS Modeler automatisch ausgewählt und erzeugen automatisch die Vorhersage.

3 Mathematische Optimierung

Im nächsten Schritt muss nun ein konkreter Produktionsplan ermittelt werden, der die vorhergesagten Verkaufszahlen möglichst gut erfüllt, dabei Nebenbedingungen einhält wie beschränkte Produktionskapazitäten, und gleichzeitig Zielkriterien optimiert, wie möglichst geringe Produktionskosten oder möglichst großer Umsatz.

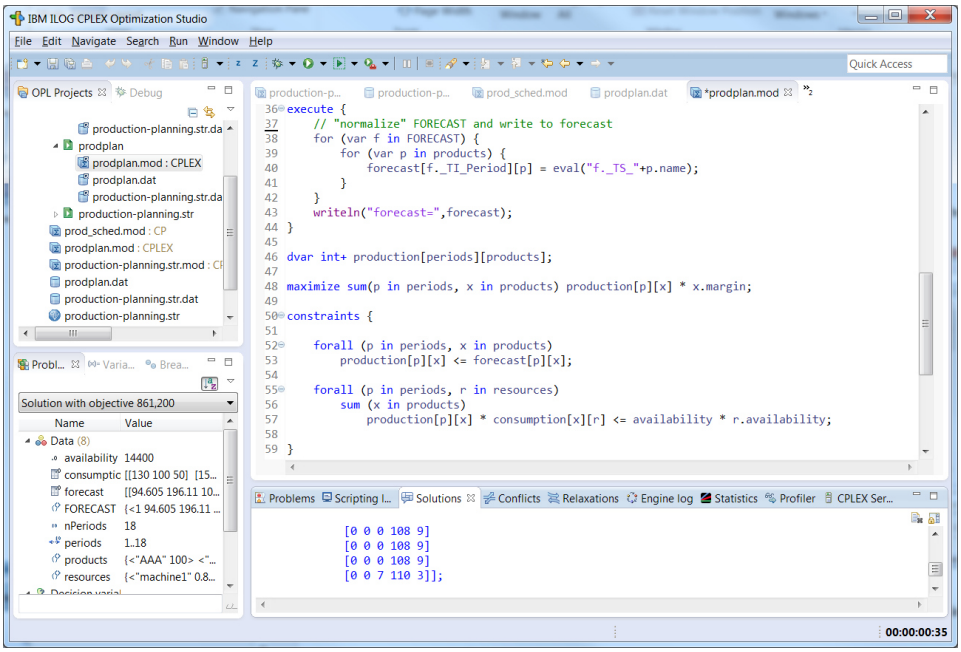


Abb. 2: Optimierungsmodell im *IBM CPLEX Optimization Studio*.

Das Optimierungsmodell wird im *IBM CPLEX Optimization Studio* modelliert: siehe Abb. 2. Dieses Modell wird dann im CPLEX Optimization Studio mit den Vorhersagedaten aus SPSS verknüpft, so dass zur Laufzeit die Vorhersage automatisch in den Produktionsplan einfließt.

4 Business Intelligence Anwendung

Produktionsplanung wird in der Realität nicht vollautomatisch von Software durchgeführt, und sei diese auch noch so intelligent. In der Praxis arbeiten Planer mit der Software und treffen am Ende die Planungsentscheidungen. Die Software unterstützt sie dabei, indem sie Planvorschläge automatisch erstellt.

Die Planungsanwendung basiert auf dem Werkzeug IBM Cognos TM1, einem Business Intelligence Werkzeug. Cognos TM1 unterstützt manuelle Planung mit einer Vielzahl intelligenter Funktionen wie automatischer Verteilung von Mengen, automatischer Aggregation und De-Aggregation, multi-dimensionalen Planwürfeln, *In-Memory* Verarbeitung usw. usf. Die oben beschriebenen Funktionen zur Vorhersage und Optimierung sind integriert in die TM1 Anwendung, siehe Abb. 3. Damit kann der Planer sowohl manuell planen, als auch interaktiv die automatische Vorhersage und Optimierung nutzen.

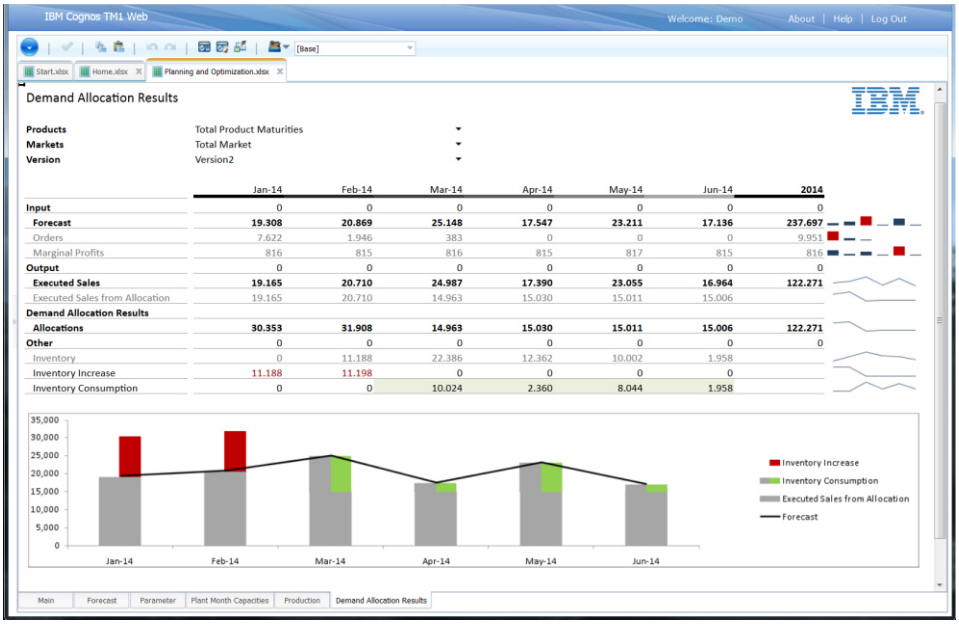


Abb. 3: Die komplette Planungsanwendung in *IBM Cognos TM1*.

5 Zusammenfassung und Bewertung

Die kombinierte Lösung aus IBM SPSS, CPLEX und TM1 erlaubt Planern sehr effizient zu planen und dabei automatische Business Analytics Methoden zu verwenden.

Ein Halbleiterhersteller, der diese Lösung vor kurzem eingeführt hat, bewertet sie folgendermaßen:

- Reduzierter manueller Planungsaufwand.
- Weniger Planungsfehler.
- Erheblich mehr Planungsszenarien können erzeigt und bewertet werden als früher.
- Generischer, konfigurierbarer Ansatz als Grundlage für weitere Planungsprozesse.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt(Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Reservierungsmanager für Cloud-Ressourcen mittels constraintbasierter Programmierung

Hans-Joachim Goltz und Armin Wolf[1]

Abstract: Es wird ein Reservierungsmanager vorgestellt, der Reservierungen von Ressourcen in einer Cloud verwaltet und eine Ressourcenplanung ermöglicht. Die Hauptaufgabe des Reservierungsmanagers besteht darin zu überprüfen, ob für jeden Dienst mit einem vertraglich vereinbarten Fertigstellungstermin die erforderlichen Ressourcen dann zur Verfügung stehen, wenn sie benötigt werden. Die Constraint Programmierung wird hierbei verwendet, um die Planungskomponente des Reservierungsmanagers zu implementieren. Der Reservierungsmanager ermöglicht eine effizientere Auslastung der Cloud-Ressourcen.

1 Einleitung

Das Ressourcenmanagement ist für das Cloud Computing ein sehr komplexes Kernproblem, das auf verschiedenen Abstraktionsniveaus betrachtet werden kann. Für einen Cloud-Provider ist es sehr schwierig, eine vorausschauende dynamische und effiziente Ressourcenzuordnung zu organisieren. In diesem Kurzbericht wird ein Manager vorgestellt, der Reservierungen von Ressourcen verwaltet und eine Ressourcenplanung ermöglicht. Für einen Cloud-Provider besteht durch die Verwendung dieses Managers die Möglichkeit, die Fertigstellung eines Dienstes zu einem bestimmten Zeitpunkt zu garantieren. Um solche zeitliche Garantien geben zu können, muss der Cloud-Provider sicher sein, dass die erforderlichen Ressourcen der Cloud in dem relevanten Zeitraum auch verfügbar sind. Durch die Verwendung des Reservierungsmanagers kann eine solche Ressourcenverfügbarkeit bereits vor der Annahme des gewünschten Dienstes überprüft werden. Außerdem kann der Cloud-Provider mittels dieses Managers auf Kundenwunsch auch garantieren, dass die Kundendaten gewünschte örtliche Regionen nicht verlassen.

Die Verwendung eines Reservierungsmanager ist nur sinnvoll, wenn für Dienste Fertigstellungszeiten garantiert werden sollen und wenn diese Dienste für eine längere Zeit einen größeren Umfang an Cloud-Ressourcen benötigen. Eine vorausschauende Reservierung von Ressourcen wird in [FBF12] unter der Anwendung eines spieltheoretischen Ansatzes diskutiert. Techniken der stochastischen Integer-Programmierung werden in [LG2010] verwendet, um ein auf SLA (Service Level Agreement) basierendes Ressourcen-Scheduling zu optimieren. Eine Scheduling-Strategie, die Reservierungen für priorisierte Jobs durchführt und ein dynamisches Scheduling realisiert, wird in [Bar13] beschrieben. In [SS12] wird ein Grundalgorithmus für eine dynamische Ressourcenzuordnung vorgeschlagen, der Vertragsvereinbarungen (SLA) und Risikoanalyse berücksichtigt. Dieser Grundalgorithmus enthält auch eine Planung von Ressourcen-Reservierungen und Rescheduling auf

[1] Fraunhofer FOKUS, 10589 Berlin, {hans-joachim.goltz, armin.wolf}@fokus.fraunhofer.de

Basis der SLA. Ein temporales Knapsack-Problem und deren Lösungsmöglichkeiten werden in [BF05] diskutiert. Ein Planungsproblem von Ressourcen-Reservierungen kann auch als ein temporales Knapsack-Problem betrachtet werden.

Für die Implementierung des hier vorgestellten Reservierungsmanagers wird softwaretechnisch die constraintbasierte Programmierung mit globalen Constraints verwendet. Die constraintbasierte Programmierung ist eine in der Praxis bewährte Technologie für die Lösung diskreter komplexer Probleme, insbesondere für das Lösen von Scheduling- und Ressourcenzuordnungsproblemen. Speziell wurde die am Fraunhofer Institut entwickelte Constraint-Solver-Bibliothek *firstCS* ([Wo12]) in die Implementierung des Reservierungsmanagers integriert.

Der beschriebene Reservierungsmanager wurde innerhalb des Projektes "EASI-CLOUDS" entwickelt, das ein Projekt des ITEA2 Programms ist (siehe auch [FL14]). Der deutsche Teil des Projektes wurde durch das Bundesministerium für Bildung und Forschung (BMBF) unter der Nummer 01 IS 11021 gefördert.

2 Reservierungsmanager

Die Hauptaufgabe des Reservierungsmanagers besteht darin zu überprüfen, ob für jeden Dienst mit vertraglich vereinbarten zeitlichen und örtlichen Bedingungen der Dienstrealisierung die erforderlichen Ressourcen dann zur Verfügung stehen, wenn sie benötigt werden. Der Reservierungsmanager befürwortet die Annahme eines Dienstes nur dann, wenn zu erwarten ist, dass die erforderlichen Ressourcen unter den gewünschten zeitlichen und örtlichen Einschränkungen auch in dem notwendigen Umfang zur Dienstrealisierung genutzt werden können. Die Planungskomponente des Managers bestimmt unter Berücksichtigung der Ressourcenverfügbarkeit den Zeitpunkt, wann ein Dienst startet und kann solche Startzeitpunkte unter Einhaltung aller Bedingungen auch verschieben.

Die Verwendung eines Reservierungsmanager ist nur sinnvoll, wenn für Dienste Fertigstellungszeiten garantiert werden sollen und wenn diese Dienste für eine längere Zeit einen größeren Umfang an Cloud-Ressourcen benötigen. Da diese Bedingungen allgemein nicht alle zu realisierende Dienste erfüllen, kann ein solcher Reservierungsmanager nur dann sinnvoll eingesetzt werden, wenn ein festgelegter Anteil der vorhandenen Ressourcen des Providers für Reservierungen zur Verfügung stehen. Der andere Anteil der Ressourcen kann dann durch Dienste ohne Reservierung und als Reserve genutzt werden.

Der Reservierungsmanager ist eng mit dem SLA-Manager verbunden. Bevor der SLA-Manager einen Kunden-Auftrag annimmt, muss er den Reservierungsmanager fragen, ob die erforderlichen Ressourcen mit den gewünschten zeitlichen und örtlichen Restriktionen zur Verfügung stehen oder unter welchen zeitlichen Bedingungen die erforderlichen Ressourcen verfügbar sind. Der Reservierungsmanager wird nur aktiviert, wenn er eine Anfrage oder Nachricht erhält. Das definierte Interface basiert auf der Kommunikation mit vorgegebenen XML-Termen. Da der Reservierungsmanager eng mit dem SLA-Manager verbunden ist und diese beiden Manager in Java implementiert sind, wurde das Interface

des Reservierungsmanagers so realisiert, dass die Kommunikation sowohl mittels XML-Termen als auch direkt innerhalb von Java erfolgen kann.

Durch eine in XML spezifizierte Ressourcenbeschreibung wird festgelegt, welche Ressourcen mit welchen Eigenschaften der Reservierungsmanager verwalten kann. Da der Reservierungsmanager nur Aussagen zur Verfügbarkeit von Ressourcen und keine Zuordnung konkreter Ressourcen realisiert, erfolgt die Ressourcenbeschreibung in einer abstrahierten Form, bei der vor allem Kapazitäten von Ressourcenklassen relevant sind. Dabei werden Ressourcen mit gleichen Eigenschaften zu Ressourcenklassen zusammengefasst. Somit kann ein Cloud-Provider auch festlegen, für welchen Teil seiner Ressourcen eine Ressourcenplanung erfolgt.

Eine Anfrage, die Verfügbarkeit von Ressourcen zu überprüfen oder Ressourcen zu reservieren, enthält Spezifikationen von kombinierten Ressourcen mit der jeweils gewünschten Anzahl. Die Spezifikation einer kombinierten Ressource definiert die zusammenhängenden Ressourcen, die für die Realisierung eines Dienstes erforderlich sind (zum Beispiel, die Ressourcen, die für eine Virtuelle Maschine benötigt werden). Eine solche Spezifikation enthält die Spezifikationen der einzelnen Ressourcen (zum Beispiel: Netzwerk, Festplattenspeicher, CPU) und Zeit-Constraints für die zeitlichen Beziehungen dieser Ressourcen untereinander. Die kombinierten Ressourcen werden von einander als unabhängig betrachtet. Wenn die gewünschten Kapazitäten der erforderlichen Ressourcen einer Anfrage in dem gewünschten Zeitintervall nicht verfügbar sind, dann kann der Reservierungsmanager nur eine Reduzierung der gewünschten Anzahl der kombinierten Ressourcen vorschlagen. Eine kombinierte Ressource ist aus Sicht des Reservierungsmanagers immer eine zusammenhängende Einheit. Die Einführung der Definition von kombinierten Ressourcen war ein wichtiges Hilfsmittel, um ein effizientes Management von Ressourcen realisieren zu können.

Bei einer Anfrage zur Reservierung von Ressourcen werden durch den Reservierungsmanager Startzeiten für jede realisierbare kombinierte Ressource zurückgegeben. Reservierungszeiten von Ressourcen können fest oder nicht fest sein. Wenn Reservierungszeiten nicht fest sind, dann können diese durch den Reservierungsmanager noch auf später unter Einhaltung aller zeitlichen Bedingungen verschoben werden. Da der Reservierungsmanager nur auf Anfragen reagiert und Zeitverschiebungen nicht automatisch mitteilt, muss ein Dienst, der noch nicht feste Reservierungszeiten erhalten hat, vor dem Start den Reservierungsmanager nach diesen Zeiten fragen.

Um zu vermeiden, dass parallel die gleichen Ressourcen für verschiedene Dienste für die gleiche Zeit reserviert werden, wurde das Reservierungsmanagement sequentiell implementiert. Zur gleichen Zeit wird immer nur die Anfrage eines Dienstes bearbeitet. Deshalb kommen alle Anfragen in eine Warteschlange und werden sequentiell bearbeitet.

3 Implementierung mittels constraintbasierter Programmierung

Die constraintbasierte Programmierung wurde für die Implementierung des Reservierungsmanagers, insbesondere für die enthaltene Planungskomponente, verwendet. Die cons-

traintbasierte Programmierung ist besonders dann geeignet, wenn komplexe Zeit- und Ressourcenplanungsprobleme deklarativ durch Entscheidungsvariablen und Randbedingungen (Constraints) modelliert werden können. Den Variablen sind jeweils Mengen von möglichen Werten zugeordnet und kennzeichnen typischerweise auch mögliche alternative Entscheidungen. Constraints repräsentieren relationale Beziehungen zwischen Variablen. Durch Constraints wird ausgedrückt, welche Bedingungen zulässige Belegungen der Variablen erfüllen müssen, und beschränkt somit die Wertekombinationen der Variablen. Eine Lösung eines Constraintproblems ist eine Zuordnung aller Variablen zu Werten aus ihren Wertebereichen, so dass alle Constraints erfüllt sind.

Für die Implementierung des vorgestellten Reservierungsmanagers wurde die am Fraunhofer Institut entwickelte Constraint-Solver-Bibliothek *firstCS* verwendet. Diese Bibliothek stellt dem Nutzer die notwendigen Konzepte zur Verfügung, um constraintbasierte Optimierungsprobleme über endlichen Wertebereichen zu modellieren und zu lösen. Der Anwendungsfokus von *firstCS* ist constraintbasiertes Scheduling und Ressourcenzuordnung. Im Folgenden wird kurz charakterisiert, wie der Reservierungsmanager mittels dieser Bibliothek realisiert wurde.

Die zeitlichen Constraints können üblicherweise durch Gleichungen und Ungleichungen modelliert werden. Globale *cumulative* Constraints werden verwendet, um Kapazitätsüberschreitungen der Ressourcen zu vermeiden. Durch ein kumulatives Constraint kann gesichert werden, dass zu keinem Zeitpunkt die gegebene Kapazität einer Ressource überschritten wird. Das *cumulative* Constraint wurde ursprünglich in [AB93] eingeführt, um Scheduling- und Platzierungsprobleme zu lösen.

Bei einer Anfrage zur Reservierung von Ressourcen für einen Dienst existieren im Allgemeinen verschiedene Möglichkeiten der Ressourcenzuordnung einer angefragten Ressourcenart. In einem solchen Fall kann die Vermeidung von Kapazitätsüberschreitung durch die Verwendung von alternativen kumulativen Constraints modelliert werden. Dies ist eine Verallgemeinerung des kumulativen Constraints bezüglich der Möglichkeit einer alternativen Ressourcenzuordnung. Ein solches Constraint ist in der verwendeten Constraint-Solver-Bibliothek *firstCS* implementiert und wurde auch für die Realisierung des Reservierungsmanagers verwendet. Die Implementierung des alternativen kumulativen Constraints in *firstCS* basiert auf einer Verallgemeinerung der in [WS05] dargestellten Ergebnisse.

Das Erzeugen von alternativen kumulativen oder kumulativen Constraints ist nur dann für eine Ressource erforderlich, wenn die Möglichkeit besteht, dass die Kapazitätsgrenzen überschritten werden. Für jede angefragte Quantität einer Ressource überprüft deswegen der Reservierungsmanager, ob für diese Ressource in dem relevanten Zeitintervall die angefragte Quantität ohne Einschränkung zur Verfügung steht. Nur wenn dies nicht der Fall ist, werden die genannten Constraints für diese Ressource erzeugt.

Für die Lösungssuche wurde eine anwendungsspezifische Strategie entwickelt, die wesentlich effizienter als die allgemeine Standardstrategie zur Lösung solcher constraintbasierter Optimierungsprobleme ist und spezifische Eigenschaften der Anwendung nutzt. Eine Anfrage zur Verfügbarkeit oder Reservierung von Ressourcen enthält die Spezifikation einer

gewünschten Anzahl von kombinierten Ressourcen. Eine kombinierte Ressource ist nur dann verfügbar (und somit reservierbar), wenn alle Ressourcen, die zu dieser gehören, auch verfügbar sind und alle zugeordneten Constraints erfüllt sind. Deswegen versucht die Planungskomponente für die spezifizierte kombinierte Ressource schrittweise konsistente Zuordnungen zu finden, bis die gewünschte Anzahl erreicht ist. Da jede dieser kombinierten Ressourcen die gleichen Ressourcen und zeitlichen Constraints enthält, wird bei der Lösungssuche Backtracking nur für die Ressourcenzuordnung innerhalb einer kombinierten Ressource erlaubt. Wenn es keine Lösung für eine kombinierte Ressource mehr gibt, dann kann die Lösungssuche abgebrochen werden und die bereits vorher zugeordnete Anzahl dieser kombinierten Ressource ist das Ergebnis. Wenn die gewünschte Anzahl von kombinierten Ressourcen nicht erreicht werden kann, dann wird versucht eine Lösung zu finden, bei der bereits reservierte kombinierte Ressourcen geplant werden, falls dies erlaubt und möglich ist. Wenn eine Anfrage verschiedene kombinierte Ressourcen enthält, dann wird dies als eine Folge von Anfragen mit gleichen kombinierten Ressourcen betrachtet, wobei die Reihenfolge durch die vom Nutzer vorgegebene Priorität bestimmt wird.

4 Schlussfolgerungen

Der vorgestellte Reservierungsmanager verwaltet Reservierungen von Ressourcen und kann durch einen Cloud-Provider verwendet werden, um die Fertigstellung von Diensten zu bestimmten Zeitpunkten garantieren zu können. Dies ist aber nur für solche Dienste sinnvoll, die für eine längere Zeit einen größeren Umfang an Cloud-Ressourcen benötigen. Um solche zeitliche Garantien geben zu können, muss der Cloud-Provider sicher sein, dass die erforderlichen Ressourcen der Cloud in dem relevanten Zeitraum auch verfügbar sind. Durch die Verwendung des Reservierungsmanagers kann eine solche Ressourcenverfügbarkeit bereits vor der Annahme des gewünschten Dienstes überprüft werden. Für die Nutzung eines solchen Reservierungsmanagers ist es notwendig, dass der Cloud-Provider einen Teil der Cloud-Ressourcen für das Reservierungsmanagement zur Verfügung stellt. Durch eine erste prototypische Implementierung des Reservierungsmanagers und deren Integration in ein Cloud-Projekt konnte nachgewiesen werden, dass das vorgestellte Konzept erfolgreich realisiert werden kann. Für diese Implementierung wurde softwaretechnisch die constraintbasierte Programmierung mit globalen Constraints verwendet.

Literaturverzeichnis

[AB93] Aggoun, A., Beldiceanu, N.: Extending CHIP in order to solve complex scheduling and placement problems. J. Mathematical and Computer Modelling 17(7), 57–73, (1993).

[BF05] Bartlett, M., Frisch, A.M., Hamadi, Y., Miguel, I., Tarim, S.A., Unsworth, C.: The Temporal Knapsack Problem and Its Solution. In: R. Bartak and M. Milano (Eds.), CPAIOR 2005, LNCS, vol. 3524, pp. 34–48, Springer, Berlin, Heidelberg (2005).

[Bar13] Barnabas, D.: Resource Reservation in the Cloud based on Infrastructure as a Service (IaaS). Research Journal of Computer Systems Engineering RJCSE, Vol. 04; Special Issue; (2013).

[FL14] Fiehe, C., Litvina, A., Tonn, J., Wu, J., Scheel, M., et al.: Building a Medical Research Cloud in the EASI-CLOUDS Project. In: Proceedings 6th International Workshop on Science Gateways, IWSG 2014, IEEE Xplore Digital library (2014).

[FBF12] Funke, D.,Brosig, F., M.Faber, M.: Towards Truthful Resource Reservation in Cloud Computing. In: Proc. of 6th International ICST Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS), pp. 253–262 (2012).

[LG2010] Li, Q., Guo, Y.: Optimization of Resource Scheduling in Cloud Computing. In: Proc. of 12th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, IEEE (2010).

[SW13] Sabharwal, N., Wali, P.: Cloud Capacity Management. Apress (part of Springer Science+Business Media), New York (2013).

[SS12] Siddesh, G.M., Srinivasa, K.G.: SLA - Driven Dynamic Resource Allocation on Clouds. In: Thilagam, P.S. et al. (Eds.), ADCONS 2011, LNCS, vol. 7135, pp. 9–18, Springer, Berlin, Heidelberg (2012).

[Wo12] Wolf, A.: firstCS - New Aspects on Combining Constraint Programming with Object-Orientation in Java. KI – Künstliche Intelligenz 26(1), 55–60 (2012).

[WS05] Wolf, A., Schlenker, H.: Realising the Alternative Resources Constraint. In: D. Seipel et al. (Eds.): INAP/WLP 2004, LNAI, vol. 3392, pp. 185–199. Springer Berlin Heidelberg (2005).

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Digitales Unternehmen – Bausteine für Effizienz, Agilität und Transparenz

Ulrich John[1]

Abstract: Digitaler Wandel, digitale Transformation, digitale Wirtschaft, digitale Agenda und ähnliche „Digitalbegrifflichkeiten" sind derzeit verstärkt Schlagwörter diverser Publikationen jeglicher Couleur und bestimmen maßgeblich die Schwerpunkte von Programmen und Projekten in Wirtschaft, Politik und Forschung. In diesem Paper skizzieren wir - ohne Anspruch auf Vollständigkeit - einige Aspekte/ Bausteine des Digitalen Unternehmens bzgl. der Digitalisierung von Unternehmen, die die Unternehmenseigenschaften Effizienz, Agilität und Transparenz erreichen helfen. Neben der Unterstützung von Planungs- und Optimierungsfunktionalitäten liegt ein besonderer Fokus auf der Schaffung zielgruppenorientierter Transparenz und auf der Informationsgewinnung für Planung- und Organisationsprozesse.

Keywords: Digitales Unternehmen, Digitale Revolution, Digitaler Wandel, Digitalisierung in Wirtschaftsunternehmen, Effiziente Prozesse, Agile Unternehmen, Transparenz

1 Einleitung

Digitalisierung scheint in diesem Jahr (2015) in vielfältigen begrifflichen Kontexten ein beherrschendes Thema in Wirtschaft, Gesellschaft und Politik zu sein. Wir beschäftigen uns in diesem Paper mit einigen Aspekten des *Digitalen Unternehmens*. Der Begriff *Digitales Unternehmen* wurde und wird mit unterschiedlichen Bedeutungen verwendet. So wurde er früher zum Beispiel oft als Synonym für Unternehmen mit „papierlosen Verwaltung- und Managementprozessen" genutzt, die durch Vermeidung von Medienbrüchen durchgängige Prozesse erlauben. In den letzten Jahren werden „rein digitale Unternehmen" wie zum Beispiel *linkedIn* und *Xing*, die ihren Umsatz und Gewinn derzeit ausschließlich durch Softwarefunktionalitäten im Internet oder in Clouds generieren, als digitale Unternehmen bezeichnet.

Wir verwenden den Begriff *Digitales Unternehmen* seit Jahren in dualer Weise. Zum einen verstehen wir unter ihm - in konsequenter „Weiterentwicklung" des Begriffes *Digitale Fabrik* (vgl. [BGW11]) - einen Oberbegriff für digitale Modelle, Verfahren und Methoden, die der ganzheitlichen Planung, Realisierung und Steuerung sowie der laufenden Verbesserung aller wesentlichen Unternehmensprozesse und Unternehmensressourcen dienen, wobei dies natürlich auch auf unternehmensübergreifende Prozesse (z.B. Supply Chains, kollaborative Produktentwicklungen in Unternehmensverbänden etc.) ausgeweitet werden kann. Zum anderen verstehen wir unter einem *digitalen*

[1] Hochschule für Wirtschaft, Technik und Kultur (HWTK), Fachbereich Wirtschaftsinformatik, Friedrichstr. 189, 10117 Berlin, ulrich.john@hwtk.de/ john@sir-john.de

Unternehmen ein Unternehmen, das einerseits über IT-basierte und softwareintegrierte Prozesse verfügt, die eine weitgehende Effizienz der Unternehmensfunktionalitäten sichern und andererseits (softwaregestützte) Prozesse und digitale Unternehmensmodelle besitzt, die eine marktgerechte Wandlungsfähigkeit/ Agilität (vgl. Abb. 1) im Sinne einer Unternehmensentwicklung beziehungsweise Unternehmenstransformation ermöglichen. Allgemeiner Konsens ist, dass Unternehmen durch Digitalisierung in diesem Sinne mittel- und langfristig Kosten sparen können, ihre Effizienz steigern und insbesondere durch die Unterstützung von agilen Unternehmenstransformationen, die rasch und kostengünstig auf neue Randbedingungen reagieren, im Wettbewerb bestehen können. Neben der reinen Kostenersparnis können Zeit- und Qualitätseffekte erzielt werden. So können beispielsweise Entwicklungszeiten für Produkte und Produktionsanlagen, für die Entwicklung und Etablierung neuartiger Services usw. verkürzt werden und die Qualität der Produkte kann deutlich erhöht werden.

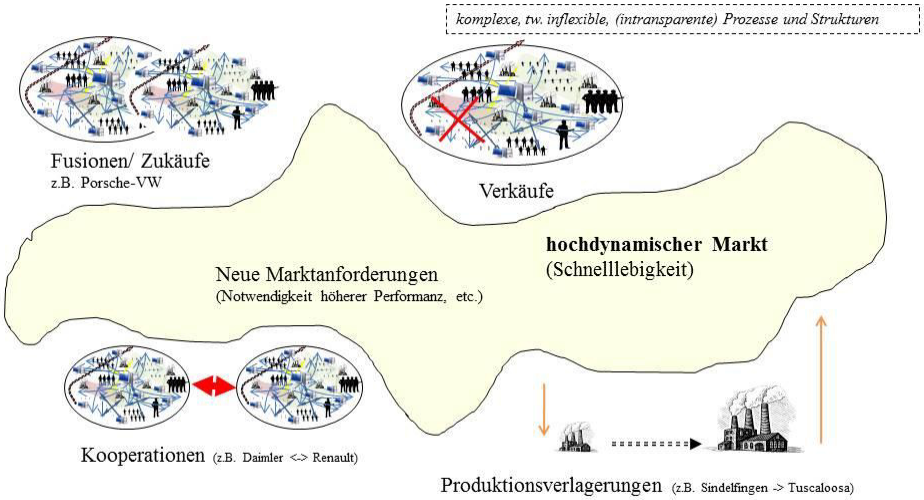


Abb. 1: Agilität - eine notwendige Eigenschaft für zukunftsfähige Unternehmen

Zusammenfassend und ergänzend formuliert, geht es neben der softwarebasierten Automatisierung und Verbesserung von Unternehmensfunktionalitäten im Wesentlichen auch um die Schaffung neuer Funktionalitäten, um Prozessoptimierungen und um Ressourcenoptimierungen. Darüber hinaus benötigt ein digitales Unternehmen Software, Modelle und Methoden für die Gewährleistung von *Flexibilität/ Agilität*, *Transparenz* und *Effizienz*. Hierfür kann sich die Wirtschaftsinformatik im Einklang mit prozessorganisatorischen Maßnahmen auf verschiedene technologische Entwicklungen der letzten Jahre auf dem Gebiet der Informatik stützen. Stellvertretend genannt seien dabei Entwicklungen auf dem Gebiet PPS-/ ERP-Systeme (siehe z.B. [Kur10]), auf dem Gebiet der deklarativen Programmierung/ Constraint Programmierung, Business Rules und Service-orientierte Architekturen (SOA), deren adäquate Anwendung entscheidende

Vorteile für die Agilität, Effizienz und Transparenz von Unternehmen mit sich bringen.

Insbesondere für große Unternehmen/ Konzerne sollte es im Kontext *Digitales Unternehmen* auch um die Schaffung von IT-basierten Prozessen gehen, die neben der Flexibilisierung und Expertenunterstützung auch die *Straffung und Vereinfachung von Managementprozessen* bewirken. Insgesamt geht es nicht nur um die Einsparung von Kosten und die Steigerung der Leistungsfähigkeit, sondern auch um die Reduzierung von vermeidbarer Verschwendung menschlicher Arbeitskraft.

In Kapitel 2 stellen wir exemplarisch für einen vernetzbaren *Baustein/ Modul digitaler Unternehmen* einen constraintbasierten Ansatz für verschiedene Planungsfunktionalitäten in einem produzierenden Unternehmen mit Fließproduktion vor. Kapitel 3 diskutiert die Wichtigkeit verlässlicher Informationen als Grundlage von manuellen oder automatisierten Managementfunktionalitäten und gibt einige Anregungen, wie die Daten-, Informations- und Wissensqualitäten insbesondere in großen komplexen Unternehmen erhöht werden können. Hierzu gehört, dass eine zielgruppengerechte Transparenz innerhalb des Unternehmens (erweitert um weitere Stakeholder, wie zum Beispiel Aktionäre) gewährleistet wird. In diesem Kontext werden betriebswirtschaftliche Planungsproblemklassen benannt, deren Lösung auf Grundlage einer hohen Transparenz und Informations-/ Wissenskorrektheit in Unternehmen durch die Services digitaler Planungskomponenten optimierend unterstützt werden kann. Das Paper schließt mit einer Zusammenfassung und einem kurzen Ausblick.

2 Operative und strategische Produktions- und Logistikplanung – Beispiel AutoMPP (Multi-Purpose Planning for Automotive)

Die Entwicklungen auf den Gebieten der Programmiersprachen, der Softwaretechnologie und die Schaffung neuer leistungsfähigerer Algorithmen für das Lösen kombinatorisch komplexer Probleme ermöglichen in den letzten Jahren verstärkt die Realisierung von leistungsstarken Planungs-, Konfigurations- und Entscheidungsunterstützungssystemen. Mit solchen Softwaresystemen lassen sich Entscheidungen fundierter treffen. Planungssysteme ermöglichen i.d.R. das Berechnen von korrekten Plänen[2] in kurzer Zeit, darüber hinaus sollen die berechneten Pläne im Sinne von Optimierungen eine höhere Güte als konventionelle Pläne besitzen. Dennoch gilt die Entwicklung guter Planungssoftware- und Entscheidungsunterstützungssysteme weiterhin als nichttrivial. Sie benötigt erfahrende Spezialisten und ist in der Regel zeit- und kostenintensiv sowie mit überdurchschnittlichen Risiken behaftet. Gelingt die Entwicklung solcher Systeme, so können die bisherigen Unternehmensprozesse in der Regel gewinnbringend reorganisiert werden, wobei die neu geschaffenen Softwaresysteme integrierte Prozessbestandteile werden. Sie leisten damit einen entscheidenden Beitrag „auf dem Weg zum digitalen Unternehmen“.

[2] D.h., die spezifizierten Randbedingungen werden erfüllt.

Im Hinblick auf die notwendige Flexibilität und Agilität von Unternehmen ist die inhärente Potenz zur Änderbarkeit und Skalierbarkeit ein weiterer Aspekt für die Entwicklung o.g. Softwaresysteme. Insbesondere der Aspekt der Anpassbarkeit kann durch die Verwendung deklarativer Spezifikationen/ Anwendung der deklarativen Programmierung erheblich begünstigt werden. Als Beispiele für leistungsfähige Planungssystemtypen im betrieblichen Umfeld kann man unter anderem Schedulingsysteme, Personaleinsatzplanungssysteme, Workflowplanungssysteme, MRO-Planungssysteme[3] oder allgemeiner *Multiressourcen-Planungssysteme* nennen.

Im Folgenden stellen wir exemplarisch überblickshaft das prototypisches Planungssystem *AutoMPP* (Multi-Purpose Planning for Automotive) vor, das ursprünglich für die Unterstützung der Produktflussplanung (*gewerkeorientierte Produktionsprogrammplanung*) in Automobilunternehmen gedacht war. Durch geeignete Erweiterungen des entwickelten Softwaremodells konnte das Anwendungsspektrum für das Planungssystem diversifiziert werden, so dass verschiedene Planungsfunktionen eines Automobilkonzerns unterstützt werden können. Die Entwicklung der prototypischen constraintbasierten Software *AutoMPP* (U. John & H. Windisch, SIR Plan GmbH[4]/ SIR Dr. John UG[5]) basiert auf jahrelangen Softwareentwicklungserfahrungen, die die beteiligten Entwickler als Mitarbeiter der damaligen DaimlerChrysler-Forschung in Forschungs- und Softwareentwicklungs-projekten sammelten. Das Planungssystem wurde in den Programmiersprachen JAVA und CHIP (vgl. www.cosytec.com) entwickelt und 2013/ 14 in Auszügen auf Basis von CHOCO (vgl. www.choco-solver.org) als ideenspendender Forschungsprototyp für einen deutschen Automobilkonzern reimplementiert.

Betrachten wir die Situation, in der sich einige Automobilkonzerne befinden bzw. in den letzten Jahren befanden: Zum einen gibt es unter den Automobil-OEMs[6] einen heftigen Wettbewerb. Neben der nachhaltigen Erschließung neuer Märkte, z.B. Russland und China, gilt es vor allem, die Produktionskosten zu senken, die Entwicklungszeiten zu reduzieren, die Produktionsdurchlaufzeiten zu senken und immer schneller immer attraktivere Modelle zu entwickeln und zu produzieren. Diese Ziele resultieren in diversen Reorganisationen, Standardisierungen und anderweitigen Aktivitäten innerhalb der Konzerne, die unseren Erachtens teilweise durch erhebliche Intransparenz und Ineffizienz geprägt sind und zusätzlich natürlich auch in Optimierungs- und Reorganisationsaktivitäten hinsichtlich der Konzern-Zuliefernetze.

Abhilfe kann neben rein prozessorganisatorischen Aktivitäten der digitale Wandel der Konzerne zu digitalen Unternehmen im Sinne der im Kapitel 1 diskutierten Kriterien bringen. Um die Möglichkeiten einer durch *AutoMPP* prototypisch motivierten Planungssoftware zu verstehen, wollen wir nun zunächst die funktionale Aufbauweise von PKW-Aufbauwerken (vgl. Abb. 2) betrachten.

[3] Maintenance, Repair and Overhaul (MRO)
[4] www.sir-plan.com
[5] www.sir-john.de
[6] Original Equipment Manufacturer (OEM)

PKW-Aufbauwerke bestehen im Wesentlichen aus drei Produktionsstufen/ Gewerken, dem Rohbau (in Abbildung exemplarisch R1, R2, R3), der Oberflächenbehandlung (Lackierung; in Abbildung exemplarisch O1 & O2) und der Montage (in Abbildung exemplarisch M1, M2 & M3). Diese Gewerke sind in der Regel in unterschiedlichen Werkshallen lokalisiert und besitzen als Produktionsuntereinheiten jeweils verschiedene Fertigungslinien, die als zentrales Bestandteil ein Fließband besitzen, das eine bestimmte Anzahl von Karosserien beherbergen kann und mit einer festgelegten – in längeren Zeiträumen konstanten - Geschwindigkeit bewegt wird. Zwischen den Rohbauhallen und den Oberflächenhallen sowie zwischen den Oberflächenhallen und den Montagehallen gibt es Karosseriepuffer, die ihrerseits wiederum aus verschiedenen Pufferelementen, z.B. Puffersträngen, Sortierern usw. bestehen. Alle Produktionseinheiten besitzen eine technische Maximalkapazität, für die einzelnen Fertigungslinien gibt es jeweils einen Schichtplan, der bestimmt, in welcher Zeit gearbeitet wird (Fließband läuft) und in welcher nicht. Zeitlich unterbrochen durch Pufferungen und Schichtpausen praktiziert ein typisches PKW-Aufbauwerk planerisch eine reine Fließfertigung: Aus den Presswerken gelieferte Blechkomponenten werden in den Rohbauhallen durch Schweißroboter zu Karosserien verschiedener *Baureihen/ Modelle* zusammengefügt und in den Rohbaupuffer transportiert.

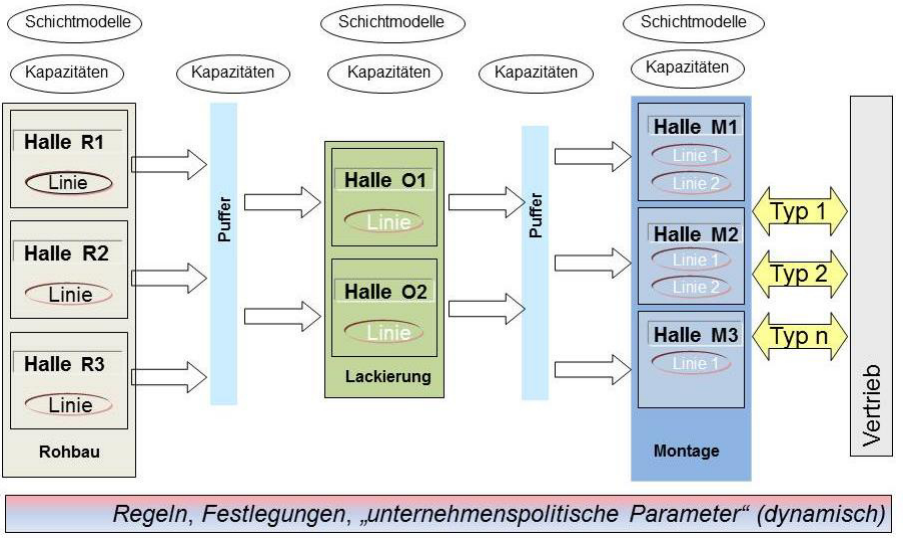


Abb. 2: PKW-Aufbauwerk

Aus dem Rohbaupuffer werden Karossen (meist sortiert in sogenannten Farbblöcken) in die Lackierung transportiert[7] und dort lackiert. Planerisch geht man hierbei in der Regel

[7] In der Praxis kann es auch sogenannte Bypass-Anlieferungen von Karosserien zu Oberflächenlinien geben, das heißt, der Puffer wird dann durch die Bypass-Karosserien nicht durchlaufen.

ebenfalls von einem jeweils einmaligen Durchfluss der Karossen aus[8]. Lackierte Karosserien werden in den Pufferstrukturen des Oberflächenpuffers gepuffert und zielgerichtet auf geeignete Montagelinien eingespeist. Da die verschiedenen Fahrzeuge in der Regel kundenindividuell ausgestattet und gefertigt werden, verwendet man hierbei das sogenannte Perlenkettenprinzip[9], damit die Logistiksteuerung die jeweils individuell zu montierenden Komponenten in der korrekten Reihenfolge bereitstellen kann. Haben die lackierten Karosserien die Montagelinien durchlaufen, so sind die jeweiligen PKWs – möglicherweise mit Ausnahme einiger durchzuführender Nacharbeiten, wie z.B. dem Anbau der Außenspiegel etc. - fertig produziert und können nach erfolgter Endabnahme dem Vertrieb und schließlich den Kunden übergeben werden.

Die Planung der Produktion ist abhängig von diversen Randbedingungen technischer und organisatorischer Natur, z.B., gibt es Restriktionen bzgl. der Zulieferungen verschiedener zu montierender Komponenten, Vorgaben bzgl. erlaubter Mischungsverhältnisse von Modellen innerhalb einer Linie, Vorgaben für Mindestpufferbefüllungen, Arbeitszeitvorgaben, die mit dem Betriebsrat teilweise bis zu zwei Jahre im Voraus abgestimmt sein müssen etc.. Die Restriktionen können klassifiziert werden und können zeitlich variabel sein. In der Regel verfügt ein Automobil-Konzern über mehrere Aufbauwerke. Grundfrage der zu behandelnden Planungsproblematik ist: Können die vom Vertrieb für bestimmte Zeitfenster angefragten Mengen verschiedener PKW-Modelle durch die Aufbauwerke des Konzerns gefertigt werden? Wenn ja, wie sind die Mengen auf die Werke aufzuteilen? Wie sind jeweilige Mengen zu terminieren? Lassen sich Zusatzschichten vermeiden? Wenn nein, dann soll die Anzahl der Zusatzschichten minimiert werden. Begrenzt ist die Anzahl der Zusatzschichten in jedem Fall durch mit dem Betriebsrat ausgehandelte Vorgaben.

Wenn die Aufteilung der Montageprogramme bzw. der sogenannten Schlussabnahmeprogramme auf die verschiedenen Werke und Montagelinien erfolgt ist, gilt es zu bestimmen, wann welche Modelle im Rohbau mit der Produktion gestartet werden müssen, damit der Fertigstellungstermin gewahrt wird, die Puffer nicht „überfahren werden“, die Sicherheitsbestände im Puffer planerisch nicht unterfahren werden, usw. Dieser Aspekt, in welcher Schicht Karossen einer Typklasse im Rohbau mit der „Fertigung beginnen“ ist planerisch wesentlich und nicht trivial. Konventionell sind derartige Berechnungen nur ungenau und sehr zeitaufwendig durchzuführen. Flexible Reaktionen auf geänderte Vorgaben und die Nachvollziehbarkeit notwendiger Produktionsstarttermine beziehungsweise auch Nachweise, dass „Produktionszahlen“/ Produktionsprogramme bei bestimmten (vorgegebenen) Randbedingungen nicht realisierbar sind, sind kaum möglich. Aus diesen Gründen ist bereits bei den skizierten planerischen Fragestellungen eine Unterstützung durch eine leistungsstarke Planungssoftware „extrem wünschenswert“ und hinsichtlich der Erreichung der Zielstellungen Effizienz, Flexibilität und Transparenz unabdingbar.

[8] In der Realität sind des Öfteren Wiederholungslackierungen notwendig. Planerisch wird hierfür eine Erhöhung der durchschnittlichen technischen Durchlaufzeit berücksichtigt.

[9] Die Einhaltung der Perlenkettenreihenfolge braucht planerisch durch AutoMPP nicht berücksichtigt werden.

Die Realisierbarkeit einer Planungssoftware, die die oben genannten Planungsaufgaben essentiell unterstützt, wurde aufbauend auf den Ideen eines früheren Prototypen (vgl. [JW06]) durch das prototypische Planungssystem AutoMPP (vgl. [JW08]) auf Basis der constraintbasierten Programmiersprache CHIP demonstriert.

Abb. 3 schematisiert die Architektur unseres Prototypen AutoMPP. Ein mit CHIP realisiertes Kernsystem konsultiert deklarative Spezifikationen der Produktionsstrukturen des Aufbauwerkes (Produktionstopologie), der bereits bekannten Randbedingungen, der Schichtmodelle sowie zusätzlich gewünschte Parametrierungen für die grafische Oberfläche des Systems, die in JAVA realisiert wurde. Aus den konsultierten Spezifikationen wird ein internes constraintbasiertes Problemlösungsmodell generiert, das aus Finite-Domain Constraintvariablen und Constraints sowie aus parametrierten Suchheuristiken besteht.

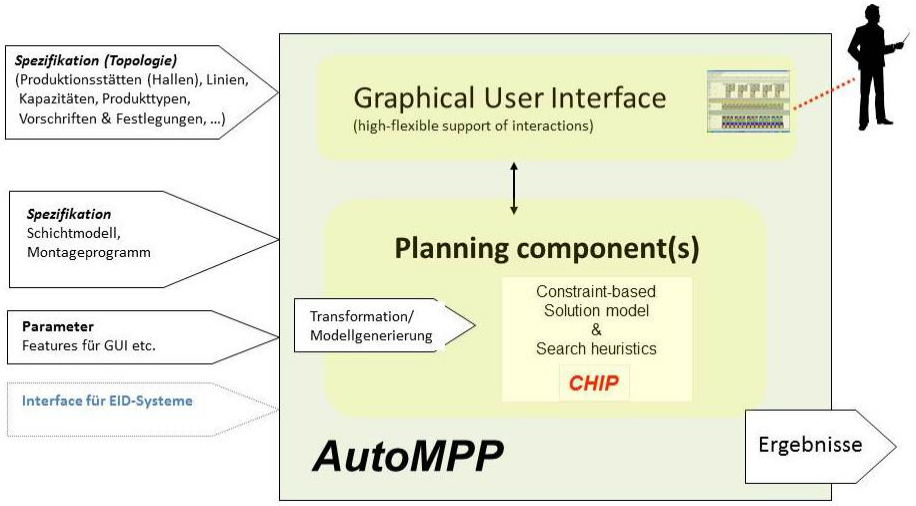


Abb. 3: AutoMPP - Architektur

Abhängig von den Intervallgrenzen der jeweiligen Schichtmodelle werden Zeitpunkte berechnet, zu denen verschiedene Systemgrößen in Form von Constraint-Variablen repräsentiert werden. Beispiele für diese Systemgrößen sind typklassenspezifische Pufferbestände, Gesamtpufferbestände, typklassenspezifische Fortschrittszahlen, Gesamtfortschrittszahlen. Für die verschiedenen Schichten werden typklassenspezifische Ein- und Ausbringungsmengen repräsentiert, desgleichen die Summen der typklassenspezifischen Ein- und Ausbringungen. Die generierten Constraintvariablen bilden die Grundlage für die durch das System zur Verfügung gestellten vielfältigen Interaktionsmöglichkeiten sowie für die Steuerung des Lösungsprozesses, der im Resultat eine Belegung aller Constraintvariablen mit gültigen Werten liefern soll und dann damit einen gültigen Gesamtplan repräsentiert.

Abb. 4 zeigt einen Snapshot des AutoMPP-Systems für eine beispielhafte Produktionstopologie in einem für die Visualisierung ausgewählten Zeithorizont[10] (eine Woche). Auf der Y-Achse des Ganttcharts werden die einzelnen Produktionsbereiche/ Fertigungslinien der spezifizierten Produktionstopologie angezeigt[11]. Diese lassen sich bei Bedarf gruppieren sowie aus- oder einblenden. Die X-Achse des Charts ist die Zeitachse.

Bereits durch die Vorgaben der für die zu planende Produktion verbindlichen Randbedingungen und des festgelegten Montageprogrammes/ Schlussabnahmeprogrammes berechnet das System erste Konsequenzen für andere Systemgrößen. Zum Beispiel sind Ober- und Untergrenzen für die verschiedenen Pufferbestände berechnet worden. In dieser Situation hat der systemnutzende Experte die Möglichkeit, interaktiv weitere Vorgaben für den zu berechnenden Plan zu tätigen. Diese können zum Beispiel auch der Abwägung verschiedener Alternativen dienen (Simulationen/ Berechnung und Bewertung von What-If-Szenarien), da die Vorgaben bei Nichtgefallen der resultierenden Konsequenzen gezielt zurück genommen werden können. Sollen keine weiteren interaktiven Vorgaben getätigt werden, dann kann eine automatische Planberechnung initiiert werden. Der berechnete Plan kann visualisiert werden und über Schnittstellen zur weiteren Verarbeitung kommuniziert werden.

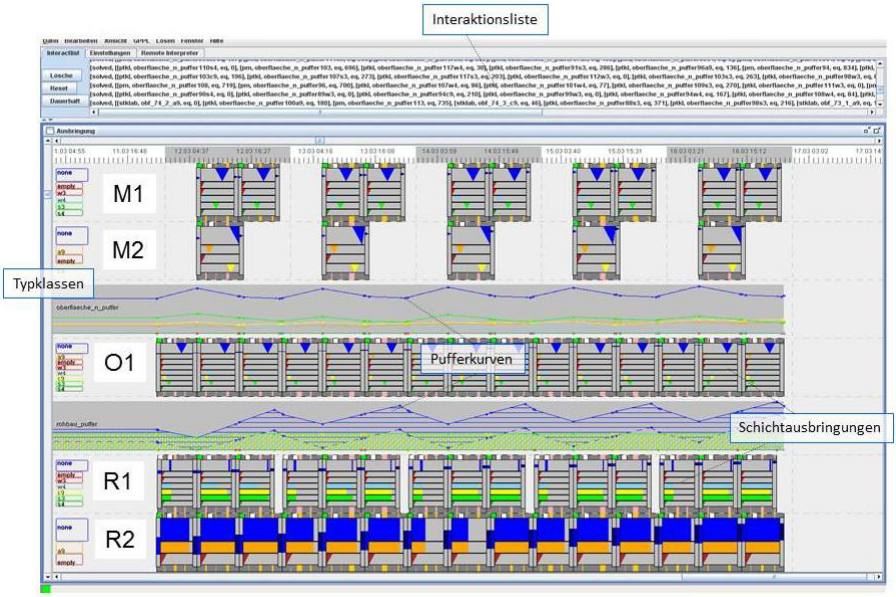


Abb. 4: GUI des AutoMPP-Prototypen - Interaktionsunterstützung

[10] Planungen können mit verschiedenen Granularitäten über wählbaren Planungshorizonten, z.B. ein Monat, ein Quartal usw. durchgeführt werden.

[11] Der Prototyp bietet die Möglichkeit, verschiedene Ausschnitte des Planungsproblems parallel zu visualisieren und als Grundlage für Interaktionen zu verwenden.

In einer weiterentwickelten Version des Prototypen hätte man nun die Möglichkeit, gezielte Umplanungen, z.B. unter Verwendung operationalisierter Constraint-Hierarchien, durchzuführen.

Aufgrund der Multidirektionalität[12] des AutoMPP zugrunde liegenden constraint-basierten Systemmodells lassen sich über die bereits behandelten Planungsfunktionalitäten hinaus verschiedene andere Planungsfunktionalitäten von PKW-Aufbauwerken (im Sinne des *digitalen Unternehmens*) unterstützen. Wesentliche Beispiele hierfür sind:

1. Produktionsanläufe (*ramp-up*) und Produktionsausläufe (*ramp-off*)

   Regelmäßig werden neue Produktmodelle entwickelt, die dann stufenweise in der Produktion „hochgefahren" werden. Da die in der Produktion beschäftigten Arbeitskräfte für die neuen Produkte/ Modelle in der Regel neue Handgriffe/ Arbeitsgänge erlernen und eintrainieren müssen, ist es für gewöhnlich nicht möglich, die Produktion des neuen Modells sofort mit „voller Stückzahl" zu starten. Stattdessen werden die neuen Modelle in der Regel mit anderen Modellen gemischt, wobei die Produktionsmengen des neuen Modells stufen-weise erhöht werden (Plateau-Fahrweise). Dieses sowie auch das „Herunter-fahren" der Produktion einer Baureihe, die nicht mehr produziert werden soll, muss im Kontext der Gesamtproduktion geplant werden, wobei sinnvollerweise verschiedene Alternativen gegeneinander abgewogen werden sollten.

2. Arbeitszeitmodellplanungen

   Ein gesonderter Funktionsbereich ist die Arbeitszeitplanung. Hierbei geht es um die Planung von für das Unternehmen sinnvollen Arbeitszeitmodellen/ Schichtmodellen. Unter anderem muss gesichert werden, dass die gesetzlichen Vorgaben und die Vorgaben des Betriebsrates eingehalten werden, dass die voraussichtlichen Produktions- und Logistikrandbedingungen erfüllt werden und so weiter. Es muss zum Beispiel geklärt werden, in welchen Produktions-einheiten Einschicht-, Zweischicht oder Dreischichtbetrieb gefahren wird und wie die Lage der Pausen innerhalb der Schichten zu wählen ist. Traditionell ist die Arbeitszeitmodellplanung von der Produktionsplanung weitgehend entkoppelt und bildet eine vorgeschaltete Planungsstufe. Durch entsprechende Modellerweiterungen kann AutoMPP zum Design von Arbeitszeitmodellen verwendet werden.

3. Arbeitskräfteplanungen

   Die Anzahl der qualifizierten Arbeitskräfte an den jeweiligen Produktionslinien hat eine direkte Auswirkung auf die „fahrbare" Geschwindigkeit der Produktionsbänder. AutoMPP beziehungsweise eine auf Basis des Prototypen entwickelbare Software kann einerseits für die Planung dieser Abhängigkeiten im operativen

[12] Die Auswertung der Constraints erfolgt ungerichtet, somit können insbesondere auch Interaktionen an nahezu beliebigen Problemgrößen in beliebiger Reihenfolge durchgeführt und ausgewertet werden.

Betrieb und andererseits für die aktive strategische Arbeitskräfteplanung/ Personalplanung eingesetzt werden.

4. strategische Produktionsplanungen und Produktionsanalagenplanungen (*Fabrikplanungen*)

   Zentrale Aufgabe der strategischen Produktions-/ Fabrikplanung ist es, Fertigungsstrukturen zu planen, die zukünftige Modellreihen in prognostizierten oder angestrebten Stückzahlen produzieren können. Zunehmend sind dabei auch Aspekte der geplanten Flexibilität von Produktionsanalagen und die Wiederverwendbarkeit von Teilen bestehender Fertigungsanlagen ausschlaggebend. Einerseits durch „händische Spezifikationsänderungen“ und deren simulative Überprüfung/ Absicherung oder aber durch entsprechende Kombination mit Konfigurationssoftwaremodellen (z.B. ConBaCon, vgl. [JG04] oder [JG03]) können solche Planungen durch AutoMPP unterstützt werden.

Das in diesem Kapitel vorgestellte prototypische System AutoMPP beziehungsweise das diesem zugrunde liegende constraintbasierte Modell stellt die solide Grundlage für einen erweiterbaren und adaptierbaren multifunktional einsetzbaren Baustein für Digitale Unternehmen dar, die Fließfertigungsproduktionen besitzen und ist exemplarisch für die auf dem Gebiet der Planungssoftwarekomponenten realisierbaren Funktionalitäten zu sehen. AutoMPP lässt sich mit anderen Bausteinen (z.B. im Sinne von Service-orientierten Architekturen) kombinieren und kann in einer weiterentwickelten Version signifikante Beiträge zum Erreichen der unternehmensbezogenen Zielstellungen Effizienz, Transparenz und Agilität liefern. Aufgrund der Deklarativität der vom System konsultierten Spezifikationen lassen sich Änderungen der Produktionstopologie (in Grenzen[13]), der vorgegebenen technischen und logistischen Randbedingungen sowie der unternehmenspolitischen Vorgaben relativ einfach umsetzen. Damit wird die Agilität des Unternehmens unterstützt. Das constraintbasierte Problemlösungsmodell unterstützt bei der Berechnung von visualisier- und auswertbaren Konsequenzen (Transparenz), die sich in Form von Kennzahlen kompakt darstellen und verwerten lassen und ermöglicht die effiziente Berechnung und Optimierung von Plänen.

Da es sich bei dem hier in Auszügen vorgestellten System um einen Prototypen handelt[14], der kombinatorisch komplexe Planungsprobleme lösen soll, sind für die Weiterentwicklung zu einem voll einsatzfähigen Softwaresystem weitere Arbeiten erforderlich. Insbesondere hinsichtlich der erforderlichen Flexibilität müssen aufgrund der Möglichkeit unterschiedlichster Merkmalskombinationen von Fertigungs- und Pufferstrukturen im Zusammenspiel mit unterschiedlichsten Arbeitszeitmodellen sowie

[13] Für den „allgemeinen Änderungsfall“ insbesondere hinsichtlich der Fertigungsstruktur ist derzeit eine Garantie der automatischen Lösbarkeit in praxisrelevanter Zeit nicht gegeben. Bei tiefer greifenden Strukturänderungen müssen strukturspezifische Heuristiken händisch angepasst werden. Bezüglich einer vollständigen Automatisierung besteht weiterer Forschungsbedarf.

[14] Auf der Basis von AutoMPP wurde in 2013/ 14 für einen deutschen Automobilkonzern ein weiterer Softwareprototyp entwickelt, der weitere Erkenntnisse hinsichtlich der Planung zukünftiger softwaregestützter Planungsfunktionalitäten liefern soll.

hinsichtlich der Forderung nach Planbarkeit für große Planungshorizonte (bis zu zwei Jahre „schichtgenau“) verschiedene Erweiterungen des Grundmodells durchgeführt und untersucht werden. Außerdem sind Untersuchungen und Entwicklungen hinsichtlich merkmalsspezifischer Suchheuristiken notwendig, damit im allgemeinen Fall (in jeglichen Kombinationen) Lösungen/ Pläne in praxisrelevanter Zeit berechnet werden können.

3 Transparenz, Informations- und Wissensqualität, Planungsfunktionalitäten für die Unternehmensorganisation

Managementprozesse und –entscheidungen betreffen unter anderem die Steuerung des laufenden Betriebs und Fragestellungen der Unternehmensorganisation beziehungsweise der Reorganisation des Unternehmens. Ein agiles Unternehmen befindet sich potenziell in einem ständigen Wandel also in fortlaufenden Reorganisationsprozessen. Neben korrektem Wissen benötigen „richtige Entscheidungen“ im Sinne der aktuellen Unternehmensziele signifikante aktuelle Daten, aktuelle Informationen und aktuelles Wissen in der richtigen Menge und in der jeweils passenden Aufbereitungsform[15]. Ein diesbezüglicher Aspekt des *Digitalen Unternehmens* ist die IT-gestützte Erfassung, Verwaltung und Auswertung vielfältiger Daten des Unternehmens und des Unternehmensumfeldes (*Data Warehouses*). Arbeiten, die sich diesbezüglich insbesondere mit großen/ riesigen Daten- bzw. Informationsmengen beschäftigen, finden sich in den Themengebieten *BIG Data*, *Databiliy* und *Smart Data* wieder.

Insbesondere großen Unternehmen und Organisationen mangelt es bis dato oft an Transparenz. Einerseits ist die Daten- und Informationslage oft schlecht, da veraltet, unvollständig oder gar fehlerhaft, anderseits gibt es unter anderem Probleme beim zeitnahen Auffinden beziehungsweise beim geeigneten Zugriff auf die gewünschten/ erforderlichen Informationen. Untermauert wird dies durch das Resultat einer Umfrage, der zufolge amerikanische Manager im Durchschnitt 25% ihrer Arbeitszeit mit dem Suchen nach Informationen verbringen (vgl. [Leh14]).

Die fehlende Transparenz behindert erheblich die Effizienz und wirkt sich negativ auf die Performanz von Unternehmen aus. Eine fehlende Transparenz bezüglich der eigenen Prozesse oder/ und der IT-Landschaft des Unternehmens führt oft dazu, dass eigentlich notwendige Reorganisationen nicht durchgeführt werden (Motto „never stop a running system“) oder extrem aufwendig und risikoreich sind[16].

Ein anderes Beispiel ist die „suboptimale Organisation“ des Gesamtmitarbeitereinsatz in großen Unternehmen. Einerseits liegen an den „entscheidenden Stellen“ oft nur unvollständige, veraltete oder „einseitige“ Informationen und Bewertungen hinsichtlich

[15] z.B. in Form managertauglicher Kennzahlen
[16] Das bestätigte sich in den vergangenen Jahren bei einigen gescheiterten Unternehmensfusionen.

der Qualifikation, der Fähigkeiten und der Erfahrungen der Mitarbeiter vor. Andererseits ist eine längerfristige Verfolgung von personellen und organisatorischen Aktivitäten, Erfolgen und Misserfolgen derzeit nicht oder nur sehr schwer möglich. Von zentraler Bedeutung sind darüber hinaus das Management des organisationalen Wissens und das gesteuerte Lernen des Unternehmens.

Oftmals sind die Prozesse, die eigentlich der Transparenzschaffung und Objektivität dienen sollen, mit anderen Prozessen und Individualzielen nicht konform. Sie wurden ursprünglich eingeführt, um die Komplexität des Unternehmens beherrschen zu können und können unseres Erachtens aufgrund des aktuellen Entwicklungsstandes der Informatik/ Wirtschaftsinformatik revolutioniert werden.

Wenn man verschiedene Großunternehmen analysiert, kann man feststellen, dass die Schaffung einer weitgehenden Transparenz (für „berechtigte Personen“) von Teilen des Managements (unausgesprochen) und auch für einige Mitarbeiter nicht gewünscht ist und somit mindestens unterschwellig behindert wird. Dies ist unter anderem durch die bestehenden Unternehmensprozesse und Regelungen bedingt. Auf detailliertere Ausführungen hierzu muss in dieser Arbeit verzichtet werden. Als unkommentierte Stichpunkte seien hier lediglich genannt: „blinder Aktionismus“ und gezielte Ausnutzung der Intransparenz, da Manager-Boni oft ausschließlich an „kurzem Zeithorizont“ bemessen werden; Konkurrenz von Cost Centern/ Werken führt zu „Geheimniskrämerei“ und verdeckten kontraproduktiven Aktivitäten; „individuelle und teilorganisationale Beschäftigungssicherung“ führt zum entwicklungshemmenden und kostensteigernden Aktionen/ Aktivitäten, managementgesteuerter „Missbrauch“ von Transparenzschaffungsinstrumenten/ -prozessen für Vergütungspolitik (Deckelung von Gehaltssteigerungen), Beförderungshemmung wegen Gefahr der Abwanderung fähiger Mitarbeiter und so weiter.

Zielgruppengerechte Transparenz - im Einklang mit der Reorganisation von Prozessen und Steuerungsinstrumenten - ist für die Performanz und Zukunftsfähigkeit eines Unternehmens essentiell. Auch bei börsenorientierten Großkonzernen könnten die meisten Aktionäre ein gesundes Interesse an einer angemessenen zielgerichteten Transparenz haben, da sie gerne gesichert haben möchten, dass ihr Kapital optimal und nachhaltig eingesetzt wird. Gleichzeitig kann eine solche Transparenzschaffung eine solide Grundlage für Agilitätseigenschaften von digitalen Unternehmen sein.

In Abb. 5 wird ein Baustein (im Sinne des *Digitalen Unternehmens*) schematisiert, der im Kontext eines Gesamtunternehmens-Wissensmanagementkonzeptes der Transparentmachung und Transparenzsicherung dienen kann.

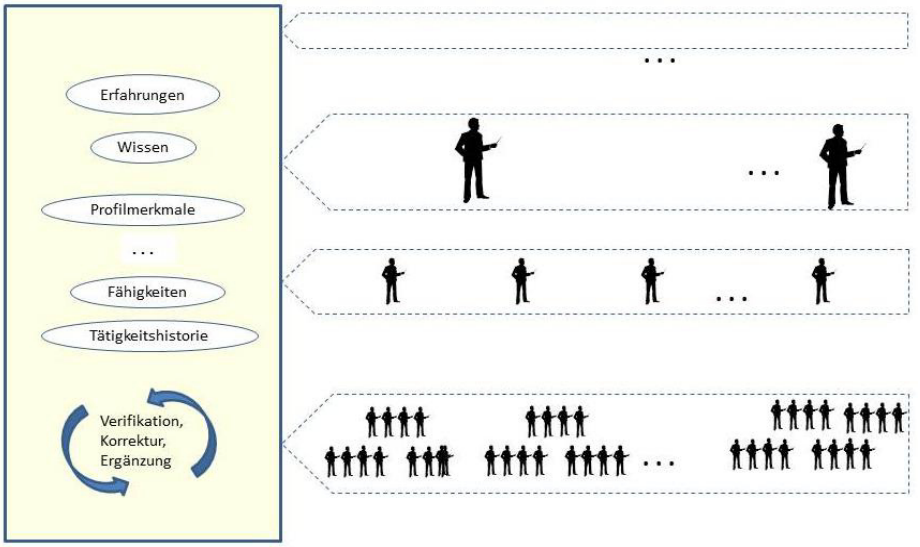


Abb. 5: Wissensmanagement in hierarchischen Organisationsstrukturen

Die Grundmerkmale des zum Beispiel über ein unternehmensweites Mitarbeiterportal (mit unterstützender Business-Prozess-Engine) realisierbaren Gesamtprozesses sind die prozessgesteuerte und -geprüfte Pflicht für Mitarbeiter und Manager, zu prozessbestimmten Zeitpunkten Informationen und ggf. Wissen zu erfassen und zu überarbeiten. Diese Informationen können sich unter anderem auf Erfahrungen, Qualifikationen, Projektstände[17], Projektergebnisse, Gründe für Projektverzögerungen oder für das Scheitern von Projekten, analysierte Profilmerkmale, Bewertungen[18], Eigenbewertungen und Organisationsstrukturen beziehen. Dabei wird eine Informationserhebung und –pflege aus multiplen Perspektiven beabsichtigt, wobei zu überlegen ist, ob auch teilanonymisierte Informationen erhoben werden sollten.

Grundsätzlich könnten zusätzlich automatisch erfasste Daten, zum Beispiel auch aus unternehmensinternen *Social Computing Systemen*, im Sinne von BIG Data hinzugefügt und aufbereitet werden.

Treten deutliche Abweichungen zwischen verschiedenen Sichten auf gleiche „Dinge“ auf, sind prozessgesteuert Verifikations- und ggf. Korrektur- und Ergänzungszyklen zu durchlaufen. Dabei sind Änderungen und Korrekturen im System in „verarbeitbarer Form“ zu begründen und zu dokumentieren. Klar ist, dass die skizzierte Vorgehensweise erstmal zusätzlichen Arbeitsaufwand erfordert. Dieser zahlt sich, unserer Meinung nach,

[17] Teile dieser Informationen können/ sollen natürlich automatisch aus anderen Systemen (ebenfalls Bausteine des digitalen Unternehmens), z.B. aus Projektmanagementsystemen, übernommen werden, die im betrieblichen Alltag verwendet werden.

[18] Auf die Diskussion von möglichen Beschränkungen durch Betriebsrat etc. sein an dieser Stelle verzichtet, da zunächst das technisch Machbare und die erreichbaren Effekte im Fokus stehen.

im Laufe der Zeit aus, da eine (für berechtigte Personen) weitgehende Transparenz erzeugt wird. Werden zum Beispiel sehr viele Projektampeln innerhalb von zwei Tagen „von *rot* auf *grün* oder *gelb* gesetzt", bevor einem Vorstandsmitglied oder dem Aufsichtsrat das Gesamtprojekt vorgestellt wird, können berechtigte Personen dies sofort erkennen und ggf. hinterfragen.

Aufgrund der über die Zeit gesammelten Informationen können automatisch „höherwertige Informationen" und Wissen deduziert werden. Beispielsweise können potenzielle Personen-Cliquen identifiziert werden. Abhängigkeiten von Personen und Cliquen zu anderen Unternehmensgrößen (z.B. nachgelagertes Scheitern von Projekten, Häufung spontaner Projektampelumsetzungen, Krankenstand der untergebenen Mitarbeiter, Fluktuation, Anzahl weiter verfolgbarer durch die gemanagte Organisationseinheit generierter Ideen etc.) können erkannt werden. Neben hier nicht weiter beschriebenen Funktionalitäten des *Wissensmanagements* und der *Planung des organisationalen Wissenserwerbs* kann der skizzierte Ansatz/ Baustein als eine unterstützende Grundlage für die *Planung und Umplanung von Organisationsstrukturen* im Sinne von agilen Anpassungen, *Weiterentwicklungen und Optimierungen von Organisationsstrukturen* genutzt werden. Dies basiert auf erforderlichen Qualifikations- und Fähigkeitskomponenten von Mitarbeiter- und Organisationseinheitsprofilen, auf Kompetenzen, auf Kapazitäten und auf geschätzten Bedarfen und Bedarfsentwicklungen. Hiermit stellt diese Frage im Grunde ein Konfigurations- oder Designproblem dar und kann jeweils durch geeignete Unterstützung digitaler Bausteine für Konfigurationen und Rekonfigurationen (siehe [JG03, JG04]) gelöst werden. Eine ähnliche Verwendung ist für die *unternehmensweite Besetzungsplanung von Projektteams* möglich. Neben den erforderlichen fachlichen Qualitäten spielen dabei im Idealfall gesammelte Erfahrungen und die „Passung und Ergänzung von Persönlichkeitsprofilen[19]" eine Rolle. Erweitert betrachtet, können Planungsfunktionalitäten bezüglich des *Multiprojektmanagements*, der *Priorisierung von Projekten* und bezüglich des *Programmmanagements* unterstützt werden. Zwei weitere durch den Ansatz unterstützbare Funktionalitäten sind die *strategische Personalbeschaffungs-* und die *Weiterqualifizierungsplanung*.

4 Zusammenfassung und Ausblick

Die meisten Unternehmen stehen vor großen Herausforderungen. Die technologischen Entwicklungen auf dem Gebiet der Informatik, die hohe Marktdynamik und der starke Wettbewerb zwingen die Unternehmen zur weitgehenden Digitalisierung. Neben der Erzeugung marktgerechter Produkte oder/ und dem Angebot von geeigneten Serviceleistungen sind *Effizienz*, *Agilität* und *gezielte Transparenz* entscheidende Wettbewerbsfaktoren. Unternehmen mit einem hohen Digitalisierungsgrad (im Sinne einer ange-

[19] Diese können durch psychologische Tests erkannt werden. Die Zulässigkeit der Erhebung, Erfassung und Verarbeitung individueller Persönlichkeitsprofile ist in jedem Falle diskussionswürdig und sollte unserer Meinung nach nur bei vorliegendem Einverständnis der jeweiligen Mitarbeiter/ -innen im Sinne der Identifikation von Win-Win Potenzialen praktiziert werden.

strebten Maximalautomatisierung) und der Fähigkeit, ihre Prozesse – einschließlich der prozessunterstützenden IT-Systeme - agil zu wandeln, bezeichnen wir als *Digitale Unternehmen*. In der vorliegenden Arbeit wurden diesbezüglich verschiedene Aspekte diskutiert. Vorgestellt wurden exemplarisch zwei Ansätze, die als Bausteine digitaler Unternehmen verwendet beziehungsweise ausgebaut werden können. Wie im Paper thematisiert, gibt es Einsatzfelder für weitere Bausteine von digitalen Unternehmen, z.B. Komponenten für Multi-Ressourcenplanung, für das Job-Shop Scheduling und für Smart-Data Funktionalitäten. Auf dem Weg zum *Digitalen Unternehmen* gilt es nun, diese Bausteine weiterzuentwickeln und zu ergänzen, damit sie z.B. als Komponenten von Service-orientierten Architekturen für vielfältige Unternehmenskonfigurationen einsetzbar sind. Hierfür sind bezogen auf constraintbasierte Komponenten unter anderem Untersuchungen und Forschungsaktivitäten hinsichtlich flexibel einsetzbarer Suchheuristiken und der dynamischer Generierung von lösungsbeschleunigenden redundanten Constraints notwendig. Ein anderes Forschungs-/ Tätigkeitsfeld bleibt die deklarative Spezifikation von Unternehmensprozessen und softwareinternen Logiken. Neben diesen Informatikaspekten gilt es, den „Weg zum Digitalen Unternehmen“ durch zu erarbeitende Vorgehensempfehlungen zu unterstützen.

Literaturverzeichnis

[BGW11] Bracht, U.; Geckler, D.; Wenzel, S.: Digitale Fabrik: Methoden und Praxisbeispiele. Springer-Verlag, 2011.

[JG03] John, U., Geske, U.: Constraint-based Configuration of Large Systems. In “Web Knowledge Management and Decision Support. (O. Bartenstein et al (Ed.))”. Lecture Notes in Computer Science, Vol. 2543, Springer-Verlag, 2003.

[JG04] John, U., Geske, U.: Integrating Time Constraints into Constraint-Based Configuration Models. In “Applications of Declarative Programming and Knowledge Management. (D. Seipel et al (Ed.))”, Post-Conference Proceedings of INAP / WLP 2004, Lecture Notes in Computer Science, Vol. 3392, Springer-Verlag, 2004.

[JW06] John, U., Windisch, H.: Factory wide production-program planning – the project gPPL. CHIP User’s Club Meeting 2006. Massy, April 2006.

[JW08] John, U.; Windisch, H.: Agility and Cost Optimization – multi-purpose planning software for automotive industry. CHIP User’s Club Meeting 2008. Paris 2008.

[Kur10] Kurbel, K.: Enterprise Resource Planning und Supply Chain Management in der Industrie. Oldenbourg Wissenschaftsverlag, 2010.

[Leh14] Lehner, F.: Wissensmanagement. Hanser-Verlag, 2014.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Constraint-basierte Planung und Optimierung von Prüfungsterminen mithilfe einer graphischen Benutzeroberfläche

Thomas Kuhnke[1] und Tom Hausten[2]

Abstract: In zwei Bachelorarbeiten ([Ha13], [Ku13]) wurde die constraint-basierte Prüfungsplanung mittels einer dafür erstellten graphischen Benutzeroberfläche umgesetzt. Dabei sollte mit der graphischen Benutzeroberfläche eine manuelle Planung von Prüfungen sowie das Anstoßen einer automatisierten Planung ermöglicht werden. Die automatisierte Planung sollte hierbei feststehende und von dem Anwender individuell festgelegte Bedingungen berücksichtigen und den Prüfungsplan möglichst nach diesen Kriterien optimieren. Bei der Planung sollten alle an der BTU Cottbus – Senftenberg angebotenen Veranstaltungen und Prüfungsordnungen berücksichtigt werden. Um die Komplexität der Planung zu reduzieren und eine bessere Übersichtlichkeit zu gewährleisten wurde die Planung in die drei Teilbereiche Tages-, Block- und Raumplanung unterteilt.

Keywords: Choco, Constraint-Programmierung, GUI, Prüfungsplanung, Vaadin

1 Einleitung

Planungsprobleme sind Probleme, die jeder Person in unterschiedlicher Ausprägung und Schwierigkeit begegnen können. Sie können im privaten Bereich aber auch im unternehmerischen Bereich auftreten. Planen allgemein ist das Erzeugen einer zeitlich geordneten Folge von Aktionen, die von der Ausgangssituation in die Zielsituation führt. Es gibt verschiedene Arten von Planungen. Die Bachelorarbeiten beschäftigen sich mit dem Scheduling. Scheduling ist ein Entscheidungsprozess zum Verteilen von Aufgaben und Tätigkeiten über einen gewissen Zeitraum mit begrenzten Ressourcen, welcher nach bestimmten Kriterien optimiert wird. Unter Ressourcen können beispielsweise die Besprechungsräume eines Unternehmens verstanden werden, die nur in gewisser Anzahl vorhanden sind. In diesem Fall könnte ein Ziel der Planung sein, maximal so viele Besprechungen gleichzeitig durchzuführen, wie es Besprechungsräume gibt [Pi12].

Bei der Prüfungsplanung an der BTU Cottbus – Senftenberg handelt es sich um ein komplexes Zeitplanungsproblem, welches derzeit in einem langwierigen Prozess manuell von den Mitarbeitern der Universität erstellt wird. Mehrere 100 Prüfungen müssen auf zwei Prüfungszeiträume à zwei Wochen verteilt werden. Dabei wird primär darauf geachtet, dass möglichst kein Student zwei Prüfungen an einem Tag schreiben

[1] BTU Cottbus – Senftenberg, Platz der Deutschen Einheit 1, 03046 Cottbus, thomas.kuhnke@b-tu.de
[2] BTU Cottbus – Senftenberg, Platz der Deutschen Einheit 1, 03046 Cottbus, tom.hausten@b-tu.de

muss und die Prüfungen möglichst zwischen Montag und Freitag, im Ausnamefall auch Samstag, stattfinden. Somit stehen 24 Tage für die Prüfungsplanung zur Verfügung. Das Problem der Planung liegt hier in der eingeschränkten Ressourcenverfügbarkeit von Zeit- und Prüfungsräumen, in denen Prüfungen durchgeführt werden können. Zudem müssen bestimmte Bedingungen beachtet werden. Beispielhaft ist dafür zu nennen, dass Professoren bzw. Studenten nicht zwei Prüfungen gleichzeitig haben dürfen. Es gibt keine feste Prüfungsdauer. In der Regel dauert eine Prüfung 90 Minuten, jedoch können diese auch 60 oder 180 Minuten dauern. Das ist abhängig von der Vorgabe der einzelnen Professoren und deren Lehrstuhl. An der BTU Cottbus – Senftenberg gibt es über 50 Prüfungsordnungen und über 1000 Veranstaltungen, die bei der Planung in unterschiedlicher Art und Weise berücksichtigt werden müssen.

Aufgrund der großen Anzahl an Informationen und Abhängigkeiten ist es sehr schwer und zeitaufwändig eine funktionierende Planung zu erstellen. Zwar ist es möglich sich an vorherigen Lösungen zu orientieren, aber trotzdem bleibt der Prozess aufwändig und fehleranfällig. Aus diesem Grund sollte ein automatisiertes Verfahren entwickelt werden, um die Planungen zu vereinfachen. Dazu gehört die automatische Prüfungsplanerstellung auf Basis der constraint-basierten Programmierung durch ein Tool sowie das Anpassen bzw. die manuelle Erstellung eines Prüfungsplans über eine graphische Benutzeroberfläche.

2 Die constraint-basierte Planung und Optimierung von Prüfungsterminen

Eine initiale Variante eines Planungstools hat die Prüfungsplanung als ein gesamtes Problem behandelt und nach einer Gesamtlösung gesucht. Dabei wurden ca. 80 Prüfungen aus vorhandenen CSV-Dateien eingelesen. Es wurden jedoch keine Optimierungen vorgenommen, was zu einer ungleichmäßigen Verteilung führte. Dieses Problem sollte mit dem neuen Planungstool beseitigt werden.

2.1 Umsetzung

Der objekt-orientierte Anteil des Planungstools wurde mit Java umgesetzt. Mithilfe der Constraint-Bibliothek Choco (Version 2.1.5) wurde das Konzept der constraint-basierten Programmierung durchgeführt. Dabei wurde auf einer SQL-Datenbank mit ca. 400 Prüfungen gearbeitet, welche unter anderem Informationen über teilnehmende Studiengänge und zeitliche Einordnungen in den Studiengängen beinhaltet. Um den Aufwand des Constraint-Lösers zu minimieren, wurde das Gesamtplanungsproblem in mehrere Teilprobleme aufgeteilt. Zunächst werden die Prüfungen auf die jeweiligen Tage verteilt, danach werden sie auf Zeitblöcke innerhalb der Tage verteilt und zum Schluss wird nach einer passenden Raumkonstellation gesucht, wie in Abb. 1 graphisch dargestellt wird. Dabei werden die Zwischenergebnisse gespeichert, so dass Änderungen

an den Block- bzw. Raumplanungsconstraints keine erneute Planung der vorangehenden Phase verursacht.

2.2 Ergebnis

Im Allgemeinen eignet sich die Constraint-Programmierung sehr gut für Planungsprobleme. Es hat sich gezeigt, dass die Optimierungsmethoden, welche Choco anbietet, durchaus vorteilhaft für die Planung sein können. Mithilfe der Optimierung ist es möglich die Nichteinhaltung von Constraints auf ein Minimum zu reduzieren. Allerdings hat das Planungstool auch gezeigt, dass Planungen schnell zu komplex werden können für den Constraint-Löser, was eine stark erhöhte Laufzeit zur Folge haben kann. Bei den 400 Prüfungen der BTU müssen allein bei der Tagesplanung aufgrund der paarweisen Überprüfung mehr als 1000 Constraints betrachtet werden. Die Komplexität der Planung ist daher sehr stark abhängig von der Menge der Prüfungen, Studiengängen sowie dem Prüfungszeitraum. Im Fall der Prüfungsplanung der BTU Cottbus – Senftenberg haben die Ergebnisse auch gezeigt, dass die vier Wochen Prüfungszeitraum für alle Eventualitäten nicht ausreichend sind und somit Kompromisse eingegangen werden müssen. Ein direkter Vergleich mit anderen Constraint-Lösern wie Gecode (Constraint-Bibliothek für C++) konnte im Rahmen der Bachelorarbeit nicht durchgeführt werden, da dieser eine fast vollständige Re-Implementierung zur Folge gehabt hätte.

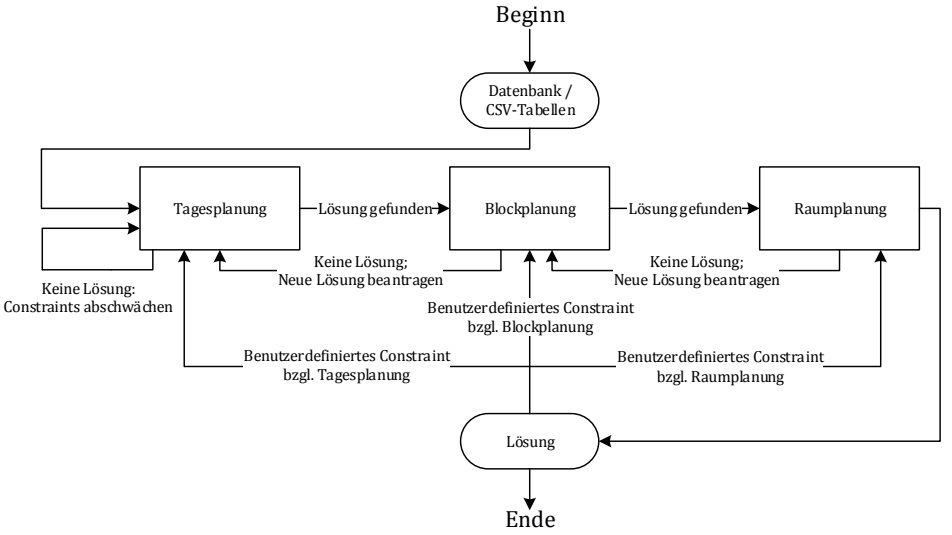


Abb. 1: Konzept zum Ablauf des Planungstools

3 Die graphische Benutzeroberfläche zur Planung und Verwaltung von Prüfungsterminen

Die graphische Benutzeroberfläche stellt einen wichtigen Teil bei der Planung von Prüfungen dar. Sie bietet dem Nutzer die Möglichkeit des manuellen Planens aller Prüfungen in den Prüfungszeiträumen und ist gleichzeitig eine Schnittstelle zum constraint-basierten Tool für die automatisierte Planung und Optimierung der Prüfungen.

3.1 Anforderungen

Die graphische Benutzeroberfläche soll zur Vereinfachungen und als Unterstützung für das Planen von Prüfungsterminen dienen. Sie soll die Möglichkeit bieten, prüfungsrelevante Daten zu bearbeiten und neue Daten hinzuzufügen. Alle relevanten Daten müssen übersichtlich und möglichst einfach dargestellt werden. Zudem sollte das Planen von Prüfungen in den drei Teilplanungen möglich sein. Jede einzelne dieser Ebenen ist frei planbar, wobei die Prüfungsplanung immer mit der Tagesplanung beginnt und mit der Raumplanung endet. Es sollte eine Schnittstelle zum constraint-basierten Planungstool geben, um Prüfungen automatisiert zu planen. Die Ergebnisse des Constraint-Lösers müssen in der Benutzeroberfläche darstellbar und zudem auch durch den Nutzer anpassbar sein. Dem Benutzer soll zudem auch die Möglichkeit gegeben sein, für einzelne Prüfungen zusätzliche Bedingungen anzugeben, welche von dem constraint-basierten Tool berücksichtigt werden. Dies kann beispielsweise ein festgelegter Termin einer Prüfung sein. Somit hat die graphische Oberfläche zur Aufgabe, das komplexe Problem möglichst einfach und übersichtlich darzustellen.

3.2 Umsetzung

Um langwierige Installationen eines Tools bei den Nutzern zu verhindern, wurde eine webbasierte Lösung angestrebt. Die Oberfläche wurde mit Framework Vaadin umgesetzt. Vaadin ist ein auf AJAX basierendes Web Application Framework, welches Ähnlichkeiten zu GWT aufweist. Anwendungen hierfür lassen sich in Java programmieren und können mit HTML und CSS verfeinert werden. Dabei läuft die komplette Logik der Anwendung auf einem Server und kann über einen Browser von den Benutzern genutzt werden. Dadurch ist das Tool plattform- und ortsunabhängig und bietet eine hohe Kompatibilität.

Alle Daten, welche für die Prüfungsplanung benötigt werden, sind in einer Datenbank hinterlegt. Die Benutzeroberfläche bietet eine tabellarische Darstellung aller zur Prüfungsplanung benötigten Daten aus der Datenbank an, in welcher die Daten bearbeitet und hinzugefügt werden können. Auch das automatische Planen der Prüfungen aus den vorbereiteten Daten aus der Datenbank wird unterstützt. In der Abb. 2 ist der Ablauf einer automatisierten Planung durch das constraint-basierte Tool aus der graphischen Benutzeroberfläche heraus dargestellt. Die fünf dargestellten Schritte

können für alle drei Teilplanungen Tages-, Block- und Raumplanung durchgeführt werden. Die Lösungen aller Planungsebenen können, wie in Abb. 3 dargestellt, in einer kalenderähnlichen Darstellung angeschaut und manuell angepasst werden. Somit bietet die Benutzeroberfläche einen hohen Planungskomfort. Die Oberfläche bietet auch die Möglichkeit, mehrere Prüfungsplanungen für einen Zeitraum zu erstellen. Zudem werden gleichzeitige Planungen mehrerer User unterstützt.

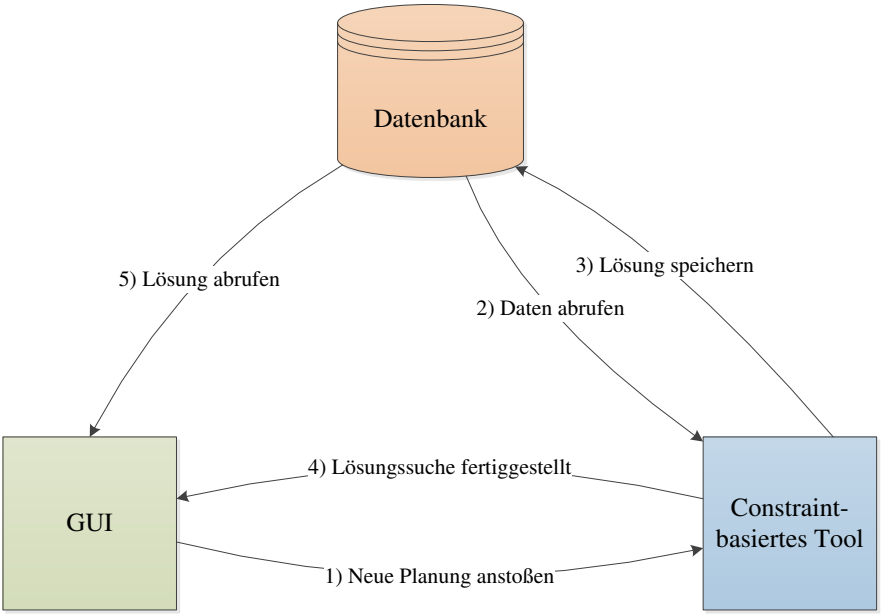


Abb. 2: Ablauf der automatisierten Planung mithilfe des constraint-basierten Tools



Abb. 3: Ausschnitt der Blockplanung

Literaturverzeichnis

[Ha13] Hausten, T.: Entwurf und Implementierung einer graphischen Benutzeroberfläche zur Planung von Prüfungsterminen. BTU Cottbus – Senftenberg, unveröffentlichte Bachelorarbeit, 2013.

[Ku13] Kuhnke, T.: Constraint-basierte Planung und Optimierung von Prüfungsterminen. BTU Cottbus – Senftenberg, unveröffentlichte Bachelorarbeit, 2013.

[Pi12] Pinedo, M.: Scheduling: Theory, Algorithms and Systems. Springer US, 2012.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Constrained Production Schedule Optimization of Output-Normalized Expenditures under Uncertainty in Shift Duration and Energy Price Forecasts

Dawn Domaschk[1] and Florin Popescu[2]

Abstract: Variability in day-ahead energy prices offers the potential for reductions in energy expenditures through flexible load planning in which production is shifted to periods of low predicted energy prices. In this paper we present a means of optimizing a production run for the following day and week given basic constraints on the start and end times of production shifts as well as the total production within a specified time frame (planning horizon). The general problem is a difficult, non-convex optimization problem over stochastic variables that can only be solved in its original form by global optimization and Monte Carlo simulation techniques. However, reasonable approximations to the original problems render it more tractable. By linearizing the energy price forecasts and the plant consumption profile, as well as discretizing the stochastic variables involved, we can reduce the planning problem to a mixed integer quadratic programming (MIQP) problem, with the customary linear bounds.

Keywords: Industrial load scheduling, variable energy pricing, day-ahead energy price forecasting.

1 Introduction

Certain energy intensive industrial processes offer the possibility of shifting production schedules (within set constraints) in order to minimize energy costs by giving precedence to low energy cost regimes. These are periods of low demand coupled with high renewable energy availability, mostly determined by weather conditions, whose forecast is uncertain. The scheduling of production volume for the upcoming day (or week, month, etc.) is further challenged by inherent variability in the actual duration and energy cost per unit produced, due to inhomogeneity in the quality of the raw material or the stochastic nature of basic sub-processes (labor, reaction dynamics, environmental conditions etc.) in a production pipeline. Many processes are discrete and serial in nature: the product is made in discrete batches/units run sequentially, meaning that variances in batch duration accumulate with time.

We present a means of optimizing a production run for the following day and week given basic constraints on the start and end times of production shifts as well as the total production within a specified time frame (horizon). The inputs to the optimization

[1] Fraunhofer Fokus, IT4Energy, Kaiserin-Augusta-Allee 31, 10589 Berlin, dawn.domaschk@fokus.fraunhofer.de
[2] Fraunhofer Fokus, IT4Energy, Kaiserin-Augusta-Allee 31, 10589 Berlin, florin,popoescu@fokus.fraunhofer.de

algorithm in our example are hourly spot market energy price data and a stochastic dynamic simulation of electric steel production. Based on the results of this analysis and taking into account the aforementioned constraints, daily and weekly batch schedules are generated, which minimize some critical aspect (e.g. expected value) of a probability distribution of the output normalized energy cost (e.g. in megawatt hours per ton of product).

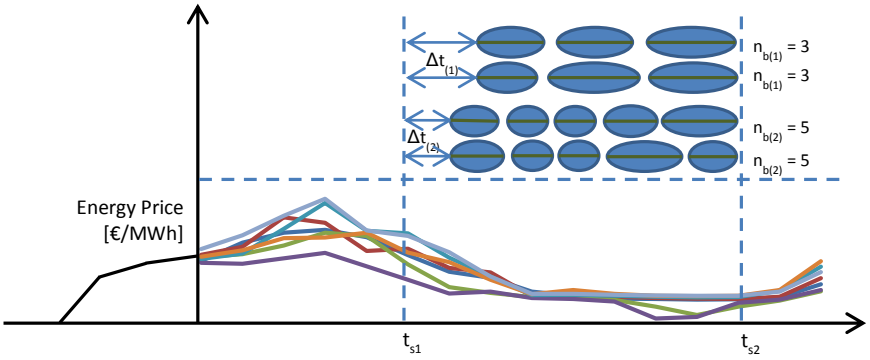


Fig. 1: Batch scheduling utilizing energy price forecast data

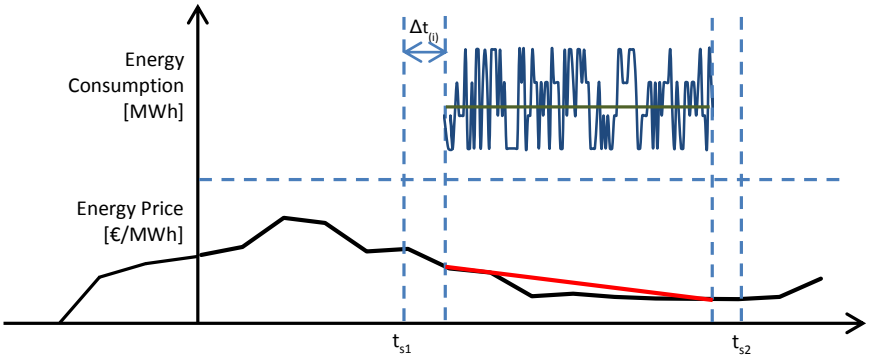


Fig. 2: Linear approximation of stochastic energy demand and spot prices

2 Approach and Constraints

The overall goal of the optimization is to minimize the sum of the cost of each shift *i* over a specified number of days *j* within the planning horizon such that the total amount of material produced within that horizon (e.g. the next week) is reached:

$$min\ E\left[\sum_{k=1}^{M}\frac{\boldsymbol{c_s}(n_{b(k)},\Delta t_k)}{n_{b(k)}}\right]\ where\ \sum_{k=1}^{M}n_{b(k)}=N \tag{1}$$

where c_s represents the total energy cost of the shift on day k, and n_b and Δt are the control parameters representing the number of batches scheduled for the particular shift and the delay before the start of the first batch in the daily production run, respectively. A production run consists of a daily serial run of batches, which are planned some hours in advance to allow for the logistics of personnel and material. To compute the cost of a shift, which is dependent on the number of batches planned during the shift and the time delay before the first batch with respect to the regularly scheduled shift start for that day, the integral of the energy cost components over the length of the shift is calculated:

$$\boldsymbol{c_s}(n_b,\Delta t)\cong\int_{s_1+\Delta t}^{s_1+\Delta \boldsymbol{s}(n_b,\Delta t)}\left(c_b(t)+\rho\boldsymbol{c_m}(t)\right)dt \tag{2}$$

where Δs is the duration of the production shift[3], c_b is the energy cost for one batch and c_m is the fixed cost of energy. The variables Δs and c_s are stochastic variables. For example, the cost at time t is equal to the base price plus paid by the consumer plus a fraction of the market price at that point in time (assuming a price structure that reflects a long-term stable price plus a current market reflecting component). Note that the fixed cost (as opposed to the variable cost) can be subtracted out for optimization purposes with no meaningful effect.

3 Optimization using Quadratic Mixed Integer Approximation

The problem described in the previous section is an optimization of scalars (the delays of the actual production runs each day) and integers (the number of batches per day) and is subject to constraints both in scalars (the end of the planned production run must arrive within the shift end) and in integers (the number of batches within the programming horizon must be preset and greater than or equal to one). We now present a means of simplifying these computations and minimizing the complexity of dynamic simulations while improving estimates of cost savings, and discuss how this basic framework is extensible to more complex situations (adaptive optimization, soft boundaries on total amount produced within the forecast horizon, etc.)

In reality, energy spot market prices vary nonlinearly with time and industrial processes may exhibit highly stochastic energy consumption. Through the linearization of energy price forecasts and the plant consumption profile, the integral of which is supra-quadratic

[3] Production shift length, $\Delta \boldsymbol{s} = n_b \cdot \Delta t_b$

within the integration limits of the optimization, as well as discretization of the stochastic variables involved, we reduce the planning problem to a MIQP approximation, bound by the duration of the production run. This approximation allows a more computationally efficient optimization of energy costs in comparison with non-linear methods. For a single forecast *i*:

$$\boldsymbol{c_{s_i}}(n_b, \Delta t) \cong \int_{t_{s_1}+\Delta t}^{t_{s_1}+\Delta t+\Delta \boldsymbol{s}(n_b,\Delta t)} \left(a_{(i)} + b_{(i)}t\right)dt \tag{3}$$

$$E(\Delta \boldsymbol{s}) \cong \sum_j P_{(j)} \Delta s_{(j)} = \sum_j P_{(j)} \Delta t_{b(j)} n_b \tag{4}$$

$$\boldsymbol{c_{s_i}}(n_b, \Delta t) \cong \mathbb{P}_2\left(a_{(i)}, b_{(i)}, \Delta t, \Delta \boldsymbol{s}\right) \tag{5}$$

Note that $\boldsymbol{\Delta t_b}$ is a stochastic variable denoting the duration of one batch; whereas, Δt is a deterministic design variable signifying the delay of a batch run relative the start of a shift. $P_{(j)}$ is a scalar which signifies the discrete probability of an event *j* whereas $\mathbb{P}_2$ signifies a second order polynomial: the coefficients are not given here for compactness but are easily derived – *n*-degree polynomials are closed under summation. Utilizing linear cost approximations and discretizing $\boldsymbol{\Delta t_b}$, the expected cost over all likely cost trajectories *i* and production run lengths *j* is calculated.

$$E(\boldsymbol{c_s}) \cong \sum_i P_{(i)} c_{s_{(i)}} \tag{6}$$

$$E(\boldsymbol{c_s}) \cong \sum_i \sum_j \mathbb{P}_{2_{(i,j)}}\left(\Delta t_{(i)}, (n_b \cdot \Delta t_b)_{(j)}\right) = \mathbb{P}_2\left(\Delta t_{b(j)}, a_{(i)}, b_{(i)}, \Delta t, n_b\right) \tag{7}$$

The constraints, apart from Eqn. 1 (linear equality constraint) are linear inequality constraints, in that the numbers of batches per day are zero or positive integers and that the end of the expected planned production run arrives within the end of the shift minus a preset safety factor. The predicted expenditures are integrations of strictly positive pricing functions and are therefore always positive; however, the method presented here does not guarantee an absolute minimum solution. In this case other general methods for global optimization can be utilized to select the greatest probabilistic minimum expenditure for each shift.

4 Conclusion and Future Work

We have presented a method for approximating a difficult, NP-hard industrial planning problem – time constrained, sequential production batch planning under uncertain (stochastic) energy costs and batch duration, such that tractable, efficient and known classes of algorithms (constrained MIQP) can be applied to unit energy cost optimization. The trade-off between the computational efficiency afforded by this approximation and the quality of the optimized solution will be quantified in future work, in which we shall compare it with a Monte-Carlo based optimization over nonlinear forecasts and dynamic simulations of the industrial process (based on our previous work), and will present preliminary results in that regard.

If price forecasts are non-linear, it follows that their integral (the batch energy expenditure) will be supra-quadratic in terms of the integration limits, which are in effect the variables for which we optimize, namely the onset and duration of the production run. Although the expenditures remain positive definite (being an integrals of strictly positive functions, namely prices), there are no guarantees of a unique local minimum, and in such case only general methods for global optimization can be used as in the relevant literature (genetic algorithms, simulated annealing, particle swarm, etc.). Over Monte Carlo evaluated cost functions, this can potentially be quite expensive from a computational standpoint. Given that forecasts are themselves uncertain, the effective performance gain of computationally expensive methods may be modest. At any rate the golden standard, against which optimization quality and the methods employed should be judged, is the expenditure, assuming perfect fore-knowledge of energy prices.

References

[HMB15] Hussain, I., Mohsin, S., Basit, A., Khan, Z., Qasim, U., Javaid, N.: A Review on Demand Response: Pricing, Optimization and Appliance Scheduling. 6th Int. Conf. on Ambient Systems, Networks and Technolgies Proceedings, Elsevier, Amsterdam, S. 1-7, 2015.

[GIO14] Grimes, D., Ifrim, G., O'Sullivan, B., Simonis, H.: Analyzing the Impact of Electricity Price Forecasting on Energy Cost-Aware Scheduling. In Sustainable Computing, S. 1-22, 2014.

[KLF12] Küster, T., Lützenberger, M., Freund, D., Albayrak, S.: Distributed Optimization of Energy Costs in Manufacturing using Multi-Agent System Technology. The Second International Conference on Smart Grids, Green Communications and IT Energy-Aware Technologies, S. 53-59, 2012.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.); INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Eine grafische, interaktive Benutzeroberfläche für den Constraint-basierten Produktkonfigurator FdConfig

Sven Löffler[1]

Produktkonfiguration gewinnt bei der ständig wachsenden Nachfrage nach Quantität und Individualität immer mehr an Bedeutung. Der *FdConfig*-Konfigurator des Lehrstuhls Programmiersprachen und Compilerbau der Brandenburgisch Technischen Universität Cottbus-Senftenberg betrachtet und löst Konfigurationsprobleme, die durch erweiterte Feature-Modelle beschrieben werden [SH11]. Erweiterte Feature-Modelle weisen eine Baumstruktur auf. Feature können dabei zu Alternativ-, Oder- bzw. Kardinalitätsgruppen zusammengefasst werden. Die Blätter des Baumes stellen Feature-Attribute dar. Jedes Blatt kann entweder ein Integer-, Enumerations- oder Bool-Attribut sein.

Der Vorteil des *FdConfig*-Konfigurators gegenüber anderen grafischen, interaktiven Konfiguratoren wie z.B. Configit oder dem CAMOS Konfigurator ist, dass *FdConfig* neben Boolschen-Attributen auch große Integer-Attribute auswerten kann.

Der *FdConfig*-Konfigurator ist Backtracking-frei, restorierbar und lässt eine beliebige Reihenfolge der Konfigurationsschritte zu, wobei der Konfigurationsraum nicht eingeschränkt wird. Bei Verwendung von komplexen Modellen ist allerdings eine Reaktionszeit von unter 250ms pro Konfigurationsschritt nicht mehr gewährleistet. Da die graphische Oberfläche nicht nebenläufig implementiert wurde, hat eine lange Berechnung zur Folge, dass der Nutzer in dieser Zeit keinerlei Steuerungsmöglichkeiten besitzt.

Im Zuge dieser Arbeit wurde der *FdConfig*-Konfigurator um eine neue interaktive graphische Oberfläche erweitert. Hauptaugenmerk lag hierbei auf der effektiven und übersichtlichen Darstellung der Integer-Attribute und einer nebenläufigen Implementierung der GUI, damit diese sich bei größeren Berechnungen des Konfigurators reaktiv verhält und nicht einfriert.

Die nebenläufige GUI Für die Darstellung des Modells wurde das Java-Toolkit *Prefuse* [Pr15] verwendet, welches es ermöglicht, Graphstrukturen graphisch darzustellen. Das Zeichnen des Graphen und die Interaktion damit wurde in einen separaten Thread ausgelagert. Dadurch wird es ermöglicht, dass konfigurationsunabhängige GUI-spezifische Einstellungen während der Berechnung eines Konfigurationsschritts weiter manipuliert werden können. Zu diesen Einstellungen, die gleichzeitig als Übersicht- und Detail-Schnittstelle dienen (overview and detail interface), gehören das Verändern der Ansicht bzw. des Ausschnittes, das Expandieren bzw. Kollabieren von Teilbäumen, das zoomen in die Grafik hinein oder aus ihr heraus, das Sichtbarmachen bzw. Verdecken

[1] Brandenburgische Technische Universität Cottbus-Senftenberg, Lehrstuhl Programmiersprachen und Compilerbau, Postfach 101344, 03013 Cottbus, Sven.Loeffler@tu-cottbus.de

von Constraint-Beziehungen und das Auswählen von Integer-Intervallen für das spätere Auswählen bzw. Eliminieren eines Integer-Attributes.

Entwurf geeigneter Visualisierungen Ein weiterer Beitrag der Arbeit war die Konzeption einer geeigneten Visualisierung für die Komponenten des Modells. Die *excludes*- und *requires*-Bedingungen zwischen Features werden durch Kanten bzw. Pfeile direkt im Graphen dargestellt. Ebenso werden Constraints durch einen Knoten, welche durch Pfeile mit den an diesem Constraint beteiligten Attributen verbunden sind, im Graphen dargestellt. Beide genannten Maßnahmen ermöglichen es dem Nutzer, leichter Zusammenhänge zwischen den Attributen abzulesen. Durch die Möglichkeit, einzelne Constraints auszublenden, können überfüllte Ansichten umgangen werden.

Interaktiven Intervallansichten Die Darstellung von Integer-Attributen wurde durch das Implementieren von interaktiven Intervallansichten kompakter gestaltet. Jedes Integer-Attribut hat eine Miniaturansicht, die dem Nutzer standardmäßig angezeigt wird (siehe Abb. 1). Durch Klick auf diese wird dem Nutzer die oberste Intervallebene angezeigt. Diese beinhaltet eine Aufteilung aller Auswahlwerte in maximal zehn Intervalle. Durch Auswahl eines der Intervalle werden dessen Unterintervalle bzw. auf der untersten Ebene dessen Werte angezeigt (siehe Abb. 1).

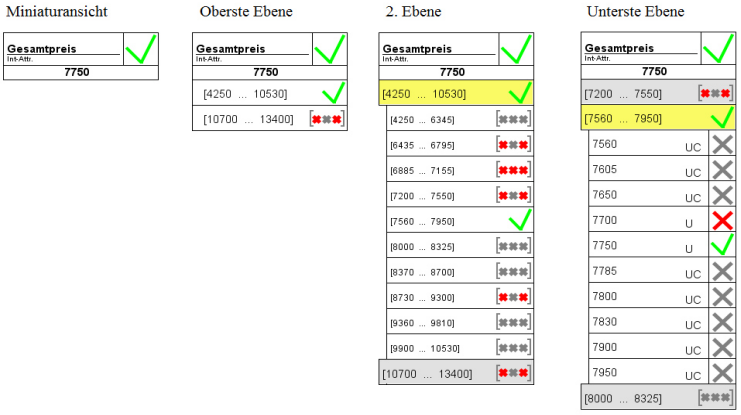


Abb. 1: Verschiedene Darstellungsformen des selben Integer-Attributes

Durch Klicken auf die grau hinterlegten Nachbarintervalle werden diese dargestellt. Ein Klick auf das gelbe Intervall lässt die Ansicht zur nächst höheren Ansicht zurückkehren. Durch ein Häkchen wird verdeutlicht, in welchem Intervall der ausgewählte Wert sich befindet, bzw. welcher der ausgewählte Wert ist. Graue Kreuze und Häkchen markierten durch den Konfigurator zur Konsistenzwahrung automatisch erzeugte Konfigurationsentscheidungen.

Sowohl mit dem Ein- und Ausblenden der verschiedenen Integer-Ansichten als auch mit dem Ein- und Ausblenden aller Knoten und Attribute des Konfigurationsbaumes wird eine Master-Detail-Beziehung unterstützt. Die verschiedenen Integer-Ansichten stellen

gleichzeitig eine sogenannte Fish-eye View dar, welche es dem Nutzer gleichzeitig ermöglicht, Ausschnitte des Konfigurationsmodells sehr detailliert und den Rest des Modells weniger detailliert darzustellen.

Die Abbildung 2 zeigt die GUI unter Verwendung eines kleinen Modells und nach Ablauf einiger Konfigurationsschritte.

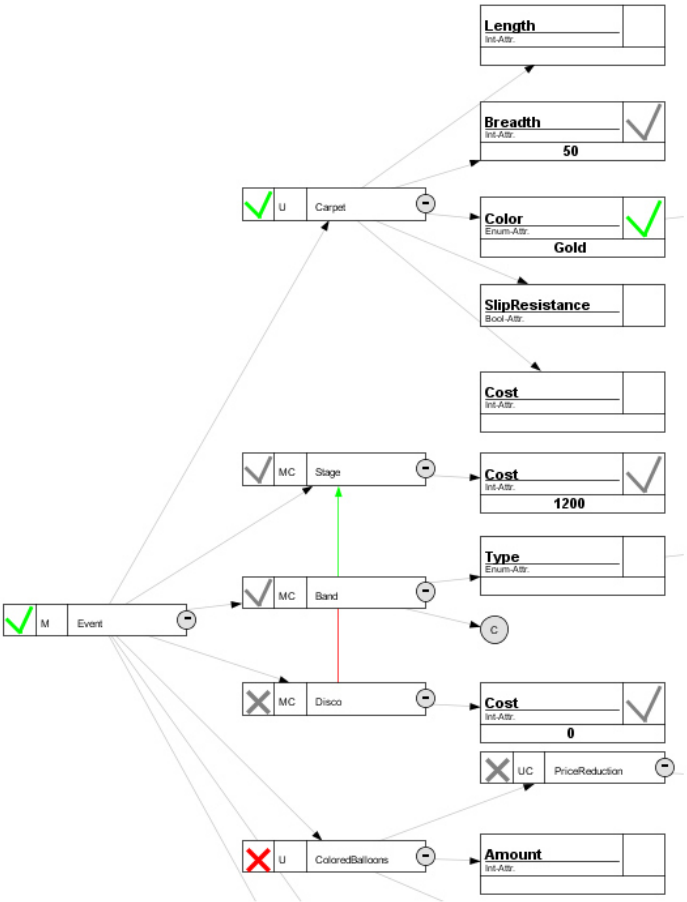


Abb. 2: Vollständige Ansicht eines Modells des *FdConfig*-Konfigurators

Die neu entstandene Visualisierung berücksichtigt bekannte HCI-Prinzipien (Human Computer Interaction) wie zum Beispiel Overview and Detail, zooming und Focus and Context. Dadurch kann die neu gestaltete Visualisierung auch Attribute mit großen Wertebereichen übersichtlich und kompakt darstellen und es dem Nutzer ermöglichen zielstrebig und ohne großen Suchaufwand bestimmte Attributwerte zu selektieren oder zu eliminieren.

[SH11] Schneeweiss, D.; Hofstedt, P. FdConfig*: A constraint-based interactive product configurator.* In: *Proceedings of 19th Conference on Applications of Declarative Programming and Knowledge Management, Wien, 2011.*

[Pr15] *prefuse – Information Visualization Toolkit,* August 2011. http://prefuse.org/.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Deklarative Modellierung und effiziente Optimierung - quo vadis? Ausgewählte Herausforderungen und Potenziale für die nächsten Jahre – zur Diskussion

Petra Hofstedt[1], Walter Hower[2], Ulrich John[3], Andreas Kmoch[4], Armin Wolf[5]

Abstract: In diesem Paper werden die für die *Brainstorming-Session* des Workshops MOC 2015 angenommenen und weiterführenden Themen aufgeführt und motiviert. Die diskutierenden Experten argumentieren, dass die Forschungsarbeit im Kontext der thematisierten – sich teilweise überschneidenden und ergänzenden – Gebiete erhebliches Potenzial für die Beherrschung großer realer Herausforderungen besitzt.

Keywords: Verteilte Constraint Programmierung in der Cloud, Optimierendes Replanning, Multi-Agentensysteme, Verteilte Scheduling- und Planungssysteme, Combinatorial Auctions, Streiks und Störungen in Verkehrs- und Versorgungsnetzen, Katastrophenmanagement, digitale Modelle von Verkehrs- und Logistiknetzen

1 Nebenläufige und Parallele Constraint Programmierung

Autorin: Petra Hofstedt

Verfahren zur nebenläufigen und parallelen Constraint-Programmierung zielen auf eine Lösung von Constraint-Problemen durch verteilte Agenten. Die Probleme werden dabei in Teilprobleme aufgeteilt, nebenläufig oder parallel gelöst und die so erzielten Teillösungen zu Gesamtlösungen kombiniert. Auf der einen Seite zählen hierzu die Konzepte der sog. verteilten Constraint-Programmierung (u.a. nach Yokoo), wobei hier i.d.R. bereits eine natürliche Verteilung der Constraint-Probleme a priori gegeben ist; ein Beispiel sind Multiagenten-Truth-Maintenance-Systeme. Demgegenüber zielen Verfahren zur echt parallelen Lösung von Constraint-Problemen ("parallel constraint processing") vor allem auf die Beschleunigung der Lösung sehr komplexer Constraint-Probleme. Bei einigen solcher Problemen ist hier eine Aufteilung probleminhärent, so dass "Sollbruchstellen" für eine Problemzerlegung einfach gefunden werden können. Dies ist z.B. bei der Dienstplanung verschiedener Personalkategorien, die letztendlich aber zusammenarbeiten und somit wenigstens teilweise aufeinander abgestimmt werden

[1] Brandenburgische Technische Universität Cottbus-Senftenberg, Lehrstuhl Programmiersprachen und Compilerbau, Postfach 101344, 03013 Cottbus, hofstedt@b-tu.de
[2] Albstadt-Sigmaringen University, Jakobstr. 6, 72458 Albstadt-Ebingen, hower@hs-albsig.de
[3] Hochschule für Wirtschaft, Technik und Kultur (HWTK), Fachbereich Wirtschaftsinformatik, Friedrichstr. 189, 10117 Berlin, ulrich.john@hwtk.de/ john@sir-john.de
[4] Toll Collect, andreas.kmoch@toll-collect.de
[5] Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin, armin.wolf@fokus.fraunhofer.de

müssen, der Fall. Teilweise ist die Zerlegung sehr komplexer Probleme aber auch nur in Planungsstufen (d.h. zeitlich) möglich, so dass man zwar aufteilen, aber nicht echt parallelisieren kann. Auch eine solche Aufteilung kann aber die Probleme u.U. reduzieren. Diskussionspunkte in diesem Themenbereich sind neben Verteilungsmodellen und einer Klassifikation von Anwendungen auch die Unterstützung der Problemzerlegung mit Userwissen unter Nutzung geeigneter Userinteraktionskonzepte.

2 Combinatorial Auctions

Autor: Walter Hower

This prominent exponential problem ([EH08]) with its huge (exhaustive search) complexity, based on the Bell number *B(n)* ([How11]), requires a heuristic view on it. [*B(n)* of a given natural number yields the # (non-empty) partitions of an *n*-set.[6]] Even after having issued several theses and student projects we still do not have a substantial guidance regarding an efficient and promising approach on the way towards the solution of the *winner determination problem* (of course without guaranteeing the optimum). Which degrees of freedom should we restrict: the number of bidders, the max. cardinality of the bundles, the # bundles, or s.th. else? Which tools, sufficiently easy to handle, do already exist? What kind of experience do other researchers have? What are your suggestions? The problem has a tremendous influence on real life, where people even electronically indicate their profiles to form purchases; it is situated in the frame of The Sveriges Riksbank Prizes in Economic Sciences in Memory of Alfred Nobel [7] in the years 1994, 1996, 2005, and 2007: John C. Harsanyi, John F. Nash Jr., and Reinhard Selten ”for their pioneering analysis of equilibria in the theory of non-cooperative games”, James A. Mirrlees and William Vickrey ”for their fundamental contributions to the economic theory of incentives under asymmetric information”, Robert J. Aumann and Thomas C. Schelling ”for having enhanced our understanding of conflict and cooperation through game-theory analysis”, and Leonid Hurwicz, Eric S. Maskin, and Roger B. Myerson ”for having laid the foundations of mechanism design theory”.

3 Replanning in großen Verkehrs- und Logistiknetzen – möglich durch Cloud-Computing?

Autoren: Ulrich John und Andreas Kmoch

Das Planen des Betriebs großer Verkehrs- und Logistiknetze beziehungsweise die Entwicklung von Planungssoftware, die solches leistet, ist nach wie vor eine große Herausforderung. Besonders in letzter Zeit erleben wir massive Störungen in Verkehrs-

[6] Thanks to Heinz Lüneburg (†) for former discussions—and teaching; this sketch is in memorial to him.
[7] not for me ☺

netzen, zum Beispiel durch Streiks im Personen- und Güterverkehr der Bahn. Ähnlich lassen sich auf dem Gebiet der Energieversorgung mit Fortschreiten der Energiewende ebenfalls Störungen befürchten. Ausgehend davon stellt sich die Frage, ob durch die technologischen Entwicklungen der letzten Jahre auf dem Gebiet des *Cloud Computings* und des *verteilten Constraint Solvings* Möglichkeiten entstanden sind, Umplanungen in großen Verkehrs- und Logistiknetzen praktikabel und optimierend durchzuführen. In diesem Kontext stellt sich auch die Frage, inwieweit es sinnvoll und möglich ist, große Gesamtnetze für Versorgung und Logistik zu digitalisieren/ deklarativ zu (re)spezifizieren, um im Falle von Katastrophen o.ä. Umplanungen und Hilfsaktionsplanungen durchführen zu können. In beiden Fällen ist die Verwendung von gesammelten Massendaten ein weiterer ergänzender Aspekt. Zum Beispiel könnte die Bahn bei der Erstellung von Notfahrplänen besonders die Strecken und Zeiträume berücksichtigen, die aufgrund gesammelter Handyeinwahldaten etc. als „besonders verkehrsstark" identifiziert werden und so weiter.

4 Cloud-Computing Framework für verteilte constraintbasierte Planung, Konfiguration und Optimierung

Autoren: Ulrich John und Armin Wolf

In den letzten Jahren ergeben sich durch die rasante Entwicklung des Cloud Computings diverse Möglichkeiten. Insbesondere bei der Verarbeitung von großen Datenmengen lassen sich neue Problemdimensionen bezwingen. In diesem Zusammenhang greifen wir zurück auf unsere in den 90er-Jahren entwickelten Konzepte für verteilte PPS-Systeme (z.B. dynamische VPPS-Systeme; vgl. [Joh95] und [GJ96]) und stellen motiviert durch das MapReduce-Framework ([DG04]) die Frage, welche softwaretechnologischen Voraussetzungen geschaffen werden müssen, um ein Cloud-basiertes Framework für die verteilte constraintbasierte Lösung von großen komplexen Planungs-, Konfigurations- und Optimierungsproblemen zu schaffen.

Literaturverzeichnis

[DG04] Dean, Jeffrey; Ghemawat, Sanjay: MapReduce - Simplified Data Processing on Large Clusters. Sixth Symposium on Operating System Design and Implementation (OSDI'04). San Francisco, CA, December, 2004.

[EH08] Egner, M; Hower, W: Computational Complexity in Constraint-based Combinatorial Auctions. In "INFORMATIK 2008 (Hegering, H.-G. et al (Ed.))", volume P-134 of Lecture Notes in Informatics, pages 541–545. GI 2008.

[GJ96] Goltz, H.-J., John, U.: Methods for Solving Practical Problems of Job-Shop Scheduling Modelled in CLP(FD). 2nd International Conference on the Practical Application of Constraint Technology (PACT'96).

[How11] Hower, W.: Diskrete Mathematik — Grundlage der Informatik. Oldenbourg Wissenschaftsverlag, München, 978-3-486-58627-5; De Gruyter, Berlin, 978-3-486-71164-6, 2010 / 2011.

[Joh95] John, U.: Nutzbarmachung großer Auftragsdatenmengen für innovative Produktionsplanungssysteme / Grundkonzepte verteilter Produktionsplanungssysteme. Arbeitspapiere der GMD Nr. 952, November 1995.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

IT-Unterstützung in Emergency Management & Response

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

IT-Unterstützung in Emergency Management & Response

Christian Erfurth[1], Jens Pottebaum[2]

Abstract: Die Informationstechnik kann in Notfallsituationen sowie bei der Vorbeugung derartig kritischer Situationen wichtige Beiträge zum Schutz und der Rettung des menschlichen Lebens liefern. IT-Unterstützung für das Notfallmanagement kann in der Verbesserung der Vorbereitung, der Abwehr und der Bewältigung von Schadenslagen helfen. Anwendungsbeispiele zeigen, dass sie z. B. die Koordination zwischen Einsatzkräften verbessern und die Lagefeststellung beschleunigen kann. IT soll in diesen Situationen helfen, komplexe und kritische Situationen zu beherrschen. Allerdings zeigen Erfahrungen, dass der Einsatz von IT als operatives Einsatzmittel oder Entscheidungsunterstützungswerkzeug auch für eine Zunahme der gefühlten Komplexität einer Einsatzlage bei vielen Beteiligten sorgen kann. Im Fokus des Workshops stehen die besonderen Herausforderungen und technischen Konsequenzen, die sich für die IT in diesem Umfeld ergeben, sowie ein Austausch über aktuelle Erkenntnisse aus der Entwicklung und der Erforschung von Lösungen. Uns interessieren dabei sowohl Ansätze, deren Fokus auf der Unterstützung bei der Rettung und Versorgung einer einzelnen Person liegt, als auch Ansätze, die sich mit einer breiten IT-Unterstützung im Notfall- und Krisenmanagement beschäftigen. Beide Aspekte können z. B. in Großschadenslagen gleichzeitig eine hohe Relevanz haben, wenn z. B. ein Rettungseinsatz durch die Beteiligung unterschiedlicher Organisationen und Behörden schnell zu einer komplexen Situation für einzelne Akteure (z. B. Notärzte) und Führungsstellen (z. B. einen Führungsstab) wird. Ein schneller Lageüberblick und ein effizienter, bedarfsgerechter Informationsaustausch zwischen den Beteiligten sind Ziele, die durch geeignete technische Lösungen und (IT-)Equipment vor, während und nach Notfallsituationen erreicht werden sollen.

1 Begutachtung und Programmkomitee

Alle eingereichten Beiträge wurden durch mindestens zwei Gutachter in ihrer Qualität und Eignung für den Workshop bewertet sowie anschließend im Organisationsteam diskutiert. Herzlichen Dank an alle, die sich an in den Begutachtungsprozess mit viel Energie eingebracht haben.

Benedikt Birkhäuser	IT.NRW
Tina Comes	University of Agder
Florian Daiber	DFKI
Wolf Engelbach	Fraunhofer IAO
Christian Erfurth	Ernst-Abbe-Fachhochschule Jena
Therese Friberg	Universität Paderborn
Rainer Koch	Universität Paderborn

[1] Ernst-Abbe-Hochschule Jena, Fachbereich Wirtschaftsingenieurwesen, Carl-Zeiss-Promenade 2, 07745 Jena, Christian.Erfurth@fh-jena.de
[2] Universität Paderborn, C.I.K., Warburger Str. 100, 33098 Paderborn, pottebaum@cik.upb.de

Birgitta König-Ries	Friedrich-Schiller-Universität Jena
Andreas Mitschele-Thiel	TU Ilmenau
Simon Nestler	Hochschule Hamm-Lippstadt
Jens Pottebaum	Universität Paderborn
Christian Reuter	Universität Siegen
Thomas Rose	Fraunhofer FIT
Volkmar Schau	Friedrich-Schiller-Universität Jena
Christian Wietfeld	TU Dortmund

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

IT-gestütztes Informationsmanagement in grenzüberschreitenden Großschadensereignissen

Thomas Pappert[1], Florian Brauner[2], Ompe Aimé Mudimu[3], Alex Lechleuthner[4], Konrad Barth[5], Andreas Lotter[6]

Abstract: Großschadensereignisse und Katastrophen stellen die beteiligten Akteure vor große Herausforderungen. Bedingt durch Faktoren wie z.B. den Klimawandel, treten diese Ereignisse immer häufiger als länderübergreifende Ereignisse auf. Informationsaustausch und deren Management sind für den Einsatzerfolg von elementarer Bedeutung. Sprachliche, semantische, rechtliche und kulturelle Unterschiede zwischen den Ländern erschweren eine strukturierte Kommunikation. Das EU-Projekt DISASTER hat aus diesem Grund eine Lösung entwickelt, den Informationsaustausch über Landesgrenzen hinweg zu verbessern.

Keywords: Großschadensereignis, Gefahrenabwehr, Informationsmanagement, Datenaustausch.

1 Einleitung

Länderübergreifende Großschadensereignisse oder Katastrophen treten in zunehmendem Maße auf und stellen die Akteure der Gefahrenabwehr vor große Herausforderungen. In einer komplexen und dynamischen Situation müssen sie in der Lage sein, ad-hoc zusammenzuarbeiten um das Einsatzziel zu erreichen. Von elementarer Bedeutung ist hierbei ein stetiger Informationsaustausch über Ländergrenzen hinweg. Dieser ist jedoch erschwert aufgrund sprachlicher, linguistischer, semantischer, rechtlicher und nicht zuletzt kultureller Unterschiede zwischen den einzelnen Gefahrenabwehrsystemen. IT-gestützte Einsatzführungssysteme (im Folgenden EMS, Emergency Management System genannt) können hierbei wertvolle Unterstützung leisten Allerdings ist der Markt für Systemlösungen sehr breit gefächert und die einzelnen Systeme stützen sich bisweilen auf unterschiedliche Formate, Protokolle und Systemarchitekturen [Ne12]. Ansätze zur Normung und Standardisierung existieren bereits, z.B. durch die DIN SPEC 91287, welche „Anforderungen an die Struktur und den Inhalt des Informationsaustausches zwischen Informationssystemen in der zivilen Gefahrenabwehr“ festlegt [Be12]. In der täglichen Arbeit stellen sich die Systeme aber nach wie vor als sehr heterogen dar. Das von der EU geförderte Forschungsprojekt DISASTER (Data Interoperability Solution At

[1] Fachhochschule Köln, Institut für Rettungsingenieurwesen und Gefahrenabwehr, Betzdorfer Str. 2, 50679 Köln, thomas.pappert@fh-koeln.de
[2] Ebd. florian.brauner@fh-koeln.de
[3] Ebd. ompe_aime.mudimu@fh-koeln.de
[4] Ebd. alex.lechleuthner@fh-koeln.de
[5] Ebd. konrad.barth@fh-koeln.de
[6] Ebd. andreas.lotter@fh-koeln.de

STakeholders Emergency Reaction) hat hierfür eine Lösung entwickelt, Informations- und Datenaustausch grenzüberschreitend zu ermöglichen.

2 Motivation

Ein Blick auf die Situation der Gefahrenabwehr in der Bundesrepublik Deutschland macht deutlich, wie komplex sich die Kooperation in Großschadensereignissen und Katastrophen darstellen kann. Die Zuständigkeit für den Katastrophenschutz in Deutschland unterliegt gemäß dem föderalistischen Staatsaufbau in Friedenszeiten den Ländern. Für die operative Gefahrenabwehr bedeutet dies, dass 16 unterschiedliche Katastrophenschutzgesetze existieren. Betrachtet man nun Feuerwehr und Rettungsdienst als elementaren Teil des Bevölkerungsschutzes, so zeigt sich auch hier eine Bandbreite von jeweils 16 unterschiedlichen Gesetzen für Feuerwehr als auch Rettungsdienst. In ihren Inhalten einander ähnlich, zeigen sich Abweichungen in den Schutzzielen sowie den jeweiligen Führungsstrukturen [HP02]. Die jeweiligen technischen Infrastrukturen in der Informations- und Kommunikationstechnik sind heterogen und unterliegen keinem bundes- oder landesweiten einheitlichen Standard. Dies zeigt sich vor allem bei den unterschiedlichen Leitstellen. Die Betreiber einer Leitstelle können aus einer Vielzahl von Anbietern von Leitstellensoftware auswählen und ihre Systemarchitektur somit an regionale Gegebenheiten und Bedürfnisse anpassen [Ne12]. Für die einzelne Leitstelle mag dieser Ansatz vorteilhaft sein, zeigt jedoch erhebliche Nachteile in der Kooperation zwischen mehreren Leitstellen. Forschungsansätze und Untersuchungen zeigen jedoch einen wachsenden Bedarf an jener Vernetzung [Lv11]. Ein intensiver Informationsaustausch ist vor allem dann von Nöten, wenn ein Großschadenereignis oder eine Katastrophe mehrere Bundesländer betrifft. Innerhalb der einzelnen Bundesländer sind dann zunächst die Hauptverwaltungsbeamten der betroffenen Kreise oder kreisfreien Städte als untere Katastrophenschutzbehörden für die Einsatzleitung verantwortlich [Fr10]. Im Rahmen der angeführten länderübergreifenden Lage müssen jedoch ad-hoc Kreise bzw. kreisfreie Städte aus n betroffenen Bundesländern zusammenarbeiten. Aufgrund unterschiedlicher Systemarchitekturen der Leitstellen und unterschiedlicher Führungsstrukturen ist die Erstellung eines einheitlichen Lagebildes eine hochkomplexe Herausforderung. Die proprietären Schnittstellen der Leitstellensoftwares lassen einen direkten Datenaustauch zwischen den Systemen oftmals nicht zu, sodass zur Erstellung des Lagebildes auf die Kommunikationskanäle Telefon, (Digital)funk oder Fax zurückgegriffen werden muss. Noch komplexer stellt sich dieser Sachverhalt dar, wenn zu den technischen Systemen der Leitstellen weitere IT-gestützte Führungssysteme hinzukommen, wie sie in Großschadenereignissen und Katastrophen mit Führung durch einen Stab üblich sind. Informationen aus dem Schadensgebiet laufen dann zunächst in den Führungsstrukturen vor Ort auf (z.B. Technische Einsatzleitung), werden dann an die Leitstellen übermittelt, müssen dort verarbeitet und schließlich an die rückwärtige Führungsebene (Führungs-/Krisenstab) weitergeleitet werden. Diese Prozesse laufen dann in allen an der Lage beteiligten Kommunen ab und müssen zu einem Gesamtbild integriert werden. Eine

immanente Gefahr des Informationsverlustes aufgrund der Vielzahl involvierter Systeme beeinflusst das Einsatzgeschehen.

Transformiert man die anhand der Bundesrepublik Deutschland dargestellte Komplexität auf Lagen, die mehrere Staaten betreffen, wird das o.g. Geflecht aus unterschiedlichen IT-gestützten Systemen noch komplexer. Zusätzlich treten weitere Probleme auf, welche unter 3 besprochen werden.

Internationale, länderübergreifende Großschadenereignisse nehmen nicht zuletzt durch den Klimawandel bedingt immer mehr zu. Als Beispiel sei hier das Hochwasser in Mitteleuropa im Sommer 2013 angeführt. Die Bekämpfung der verheerenden Auswirkungen solcher Ereignisse endet nicht an Landesgrenzen, sondern muss in einer integrierten Reaktion aller beteiligter/betroffener Länder in enger Abstimmung und Koordination stattfinden. Auf europäischer Ebene existieren hierfür Mechanismen, um eine möglichst reibungslose Zusammenarbeit zu gewährleisten. Der *EU Civil Protection Mechanism* wurde im Jahr 2001 etabliert. Ihm gehören sämtliche EU Mitgliedsstaaten sowie Island, Norwegen und die ehemalige jugoslawische Republik Mazedonien an. Als Koordinierungszentrum fungiert das *Emergency Response Coordination Centre* (ERCC). Das ERCC sammelt und analysiert rund um die Uhr Daten und Information zu EU- und weltweiten Krisen und Katastrophen [Eu15]. Zur Verfügung stehende Informationsquellen sind hierfür u.a. das *European Flood Alert System* (EFAS), das *European Forest Fire Information System* (EFFIS), *Global Disaster Alerts and Coordination System* (GDACS) oder *Meteoalarm* [Eu14]. Außerdem entwickelt das ERCC Einsatzpläne für Einsatzkräfte, Experten und Equipment und disponiert diese anhand eines laufenden Abgleichs zwischen Hilfsangeboten und Nachfragen aus den betroffenen Ländern. Zu den nationalen Kontaktstellen der Länder (in Deutschland das Gemeinsame Melde- und Lagezentrum des Bundesamtes für Bevölkerungsschutz und Katastrophenhilfe) steht das ERCC über das *Common Emergency Communication and Information System* (CECIS) in Verbindung. Hierüber können Mitteilung über Bedarfe sowie Hilfskapazitäten angemeldet und kommuniziert werden. Das ERCC koordiniert den Abgleich zwischen Hilfegesuchen und Hilfsangeboten auf zwischenstaatlicher Ebene. Kommunikation, Informations- und Datenaustausch finden zwischen den nationalen Kontaktstellen und der EU statt. Punktuell existieren bereits Initiativen zwischen benachbarten Staaten um die grenzüberschreitende Zusammenarbeit unmittelbar zu verbessern. In der Region Maas-Rhein wurde die Zusammenarbeit zwischen Feuerwehren und Rettungsdiensten aus Deutschland, Belgien und den Niederlanden durch das Projekt Emric+ erheblich verbessert [Em15]. Doch trotz solcher Initiativen stellen sich Zusammenarbeit und Kommunikation über Landesgrenzen hinweg sehr komplex und problembehaftet dar, wie im Folgenden beschrieben wird.

3 Hindernisse in der Kommunikation

Die Gefahrenabwehrsysteme der EU-Mitgliedsstaaten unterscheiden sich zum Teil

erheblich. Eine reibungslose Kommunikation wird durch sprachliche, semantische, rechtliche und kulturelle Unterschiede zwischen den einzelnen Staaten erschwert.

Sprachliche Unterschiede:

Innerhalb der 28 EU-Mitgliedsstaaten existieren 24 offizielle Amtssprachen. Wenn auch im Einsatz immer nur einige wenige Länder und deren Sprachen kooperieren müssen, kann eine Kommunikation durch die Sprachbarrieren unmöglich gemacht werden. Informationsaustausch auf einfachen technischen Wegen, wie Telefon oder E-Mail, kann ohne Hilfsmittel wie Wörterbücher, Übersetzungssoftware etc. kaum bewerkstelligt werden. Gerade in der Gefahrenabwehr ist die Sprache zudem geprägt durch eine sehr technische, fachspezifische Terminologie. Viele Begrifflichkeiten lassen sich mitunter nicht eins zu eins übersetzen. Eine angemessen Reaktion auf Großschadensereignisse ohne die Möglichkeit der direkten Kommunikation kann den Einsatzerfolg massiv gefährden, wenn Informationen durch Sprachbarrieren gar nicht oder durch Übersetzungsfehler falsch bei der Gegenstelle ankommen.

Semantische Unterschiede:

Selbst wenn eine Kommunikation mithilfe von Dolmetschern, Wörterbüchern und Übersetzungssoftware o.ä. sichergestellt wird, ist eine korrekte semantische Interpretation des Informationsgehalts noch nicht zwangsläufig sichergestellt. Die Gründe sind vielfältig. Wenn in Deutschland von einem Löschzug der Feuerwehr die Rede ist, so kann sich jeder Akteur in der Gefahrenabwehr darunter etwas vorstellen. Beispielhaft sei hier der Löschzug nach AGBF (Arbeitsgemeinschaft der Leiter der Berufsfeuerwehren) genannt: Dieser fordert 10 Funktionen innerhalb von 8 Minuten an der Einsatzstelle, nach 13 Minuten insgesamt 16 Funktionen [Ar98]. Doch wie wird in den Ländern der EU der Begriff Löschzug interpretiert? Existiert überhaupt ein Äquivalent? Diese Unklarheiten können z.B. bei der grenzüberschreitenden Nutzung von Ressourcen problematisch werden, wenn auf eine Anforderung mitunter völlig anders reagiert wird als erwartet. Als ein weiteres Beispiel sei hier ein Rettungswagen angeführt. Diese unterliegen zwar bzgl. ihrer Ausstattung der DIN EN 1789 [Be14], jedoch ist die tatsächliche Einhaltung dieser Norm keineswegs selbstverständlich und nicht verpflichtend. Neben diesem technischen Einsatzwert können auch die taktischen Einsatzwerte erheblich voneinander abweichen. Ein Beispiel: So ist z.B. in Deutschland der Einsatz eines Notarztes bei vital gefährdeten Patienten vorgesehen, in Paramedic-basierten Rettungsdienstsystemen spielen Notärzte keine Rolle. Die Kompetenzen des eingesetzten Personals sind dadurch nur sehr eingeschränkt vergleichbar und der taktische Einsatzwert eines Rettungswagens höchst uneinheitlich.

Rechtliche Unterschiede:

Die Vielzahl der in Deutschland geltenden Gesetze in der Gefahrenabwehr wurde bereits umrissen. Überträgt man diesen Zustand in einen europäischen Kontext, werden die rechtlichen Rahmenbedingungen der Gefahrenabwehr noch wesentlich komplexer. In föderalistischen Staaten, wie Deutschland und Österreich, liegen weitreichende

Gesetzgebungskompetenzen auf der Ebene der föderalen Staatsorgane. In zentralistischen Staaten, wie z.B. Frankreich, liegen die Kompetenzen hingegen auf nationaler Ebene. Die inhaltliche Ausgestaltung der Gefahrenabwehrgesetze ist EU-weit sehr unterschiedlich und somit untereinander nur schwer vergleichbar. Dies hat durchaus direkten Einfluss auf das Einsatzgeschehen. So ist in jeder Einsatzlage zu eruieren, auf welchen Ebenen eine Kommunikation stattfinden muss. Wer sind die jeweiligen Gesamtverantwortlichen, wer sind die Einsatzleiter im direkten Schadensgebiet, wer trägt politisch-administrative Verantwortung? Nur wenn diese Sachverhalte eindeutig geklärt sind, können strukturierte und zielgerichtete Kommunikation und Informationsaustausch stattfinden.

Kulturelle Unterschiede:

Unter kulturellen Unterschieden wird in diesem Kontext die Gesamtheit der Gefahrenabwehrsysteme gemeint. Darunter wird der generelle Aufbau eines nationalen Gefahrenabwehrsystems verstanden und impliziert u.a. den Einsatz von hauptamtlichen und ehrenamtlichen Kräften, Paramedic- oder Notarzt-basiertes Rettungsdienstsystem, Aufstellung von spezifischen Einheiten für den Katastrophenschutz, Organisation des Katastrophenschutzes, Einsatztaktiken, technische Ausrüstung, Führungsstrukturen etc. Außerdem lässt sich unter kulturellen Unterschieden der Einsatz von taktischen Zeichen in der Gefahrenabwehr subsumieren (vgl. u.a. [St10]). Gerade für das Führen von Lagekarten sind diese Zeichen von größter Bedeutung und erfahren auch in der Entwicklung der DISASTER Lösung große Beachtung. So müssen jene taktischen Zeichen nicht nur zwischen Lagekarten ausgetauscht und jeweils am korrekten Ort platziert werden, vielmehr müssen sie auch inhaltlich übersetzt werden. Dies bedeutet, dass z.B. das deutsche Zeichen für Rettungswagen in das jeweilige Äquivalent übersetzt werden muss.

4 Technische Entwicklung

Neben den genannten, mehrheitlich organisatorisch taktischen Unterschieden in der Gefahrenabwehr, tritt auch im internationalen Kontext die Vielfalt der Möglichkeiten des Einsatzes von Informationstechnik zu Tage. Ist die Bandbreite der eingesetzten technischen Systeme schon innerhalb Deutschlands äußerst heterogen, wird diese Problematik nochmals verschärft bei einer internationalen Betrachtung. Die Gefahrenabwehrsysteme und –strukturen der EU Mitgliedsstaaten sind jeweils aus geschichtlichem, (gesundheits)politischem, ökonomischem und kulturellem Kontext erwachsen und entwickelt worden. Dies bedingt eine kaum zu vergleichende Vielfalt der Systeme. Eine Harmonisierung ist aufgrund der autonomen Gesetzgebung innerhalb der Mitglieder des Staatenverbunds nicht möglich und gerade in der Gefahrenabwehr im Hinblick auf regionale Charakteristika auch nicht erstrebenswert. Vor diesem Hintergrund wären auch Versuche, einheitliche, standardisierte EMS in den Gefahrenabwehrsystemen zu installieren, zum Scheitern verurteilt. Auch die Schaffung

weiterer Ressourcen, z.B. eine zusätzliche Software zum länderübergreifenden Datenaustausch wäre im Sinne der ohnehin schon bestehenden Unübersichtlichkeit ein falscher Ansatz. Ziel der Forschung des DISASTER Projekts war daher, grenzüberschreitende Kommunikation zu ermöglichen, ohne zusätzliche Systeme zu entwickeln und ohne in bestehende Systeme einzugreifen. Vielmehr steht die Vernetzung von Systemen unter Beibehaltung ihrer jeweiligen Datenformate und Protokolle im Mittelpunkt der Forschung. Die entwickelte Lösung sieht einen zweifältigen Ansatz vor:

- Entwicklung eines Gateway um existierende Systeme miteinander zu vernetzen.
- Entwicklung einer Ontologie um auszutauschende Informationen zu übersetzen und zu mediieren.

4.1 DISASTER Gateway

Vor dem Hintergrund, dass EMS mit jeweils spezifischen Datenformaten und Protokollen arbeiten, muss eine Lösung gefunden werden diese Daten so zu bearbeiten, dass sie mit weiteren EMS ausgetauscht werden können. Der DISASTER Gateway besteht daher aus einem Netzwerk mehrerer Komponenten: aus dem sog. DISASTER Core sowie den Mediationskomponenten.

DISASTER Core:

Auszutauschende Daten zwischen Systemen haben wie erwähnt nicht notwendigerweise dasselbe Datenformat. Mit der Entwicklung des DISASTER Core ist ein Tool zur Lösung dieses Problems geschaffen worden. Der Core ist in der Lage, unterschiedliche Datenformate in RDF (Resource Description Framework) umzuwandeln. Der Prozess, ein erhaltenes Datenformat in RDF umzuwandeln wird hier *Lifting* genannt, während der gegenläufige Prozess, nämlich RDF in ein benötigtes Datenformat umzuwandeln, als *Lowering* bezeichnet wird. Folgende Datenformte ist der DISASTER Core in der Lage in RDF umzuwandeln:

- Geographic Markup Language (GML)
- Relational Database (RDB)
- ESRI Shapefile (SHP)
- Keyhole Markup Language (KML)
- Portable Network Graphics (PNG)
- JSON Geometry and Feature Description (GeoJSON)
- Portable Document Format (PDF)

Diese Formate wurden ausgewählt, weil es sich hierbei um geläufige und viel genutzte

Formate handelt.

Nach der Transformierung der Daten in RDF findet eine ggf. notwendige Mediation der Daten statt. Soll z.B. ein taktisches Zeichen auf einer Lagekarte ausgetauscht werden, so muss dieses Zeichen in die „Sprache“ der Gegenseite übersetzt werden und an der korrekten Position auf der Lagekarte des Empfängers platziert werden [Ca13] (siehe Beispiel unter 4.3). Die Übersetzung geschieht anhand der EMERGEL Ontologie.

DISASTER Mediator:

Als DISASTER Mediator werden die direkten Gateways teilnehmender EMS an die DISASTER Architektur bezeichnet. Der Output Mediator ermöglicht einem EMS den Export von Informationen/Daten für die Bereitstellung an den DISASTER Gateway. Der Input Mediator ermöglicht einem EMS wiederum den Import von bereitgestellten, angepassten Informationen.

Derzeit verfügen die EMS noch nicht über die geforderten Schnittstellen, jedoch wurde der DISASTER Mediator in zwei erfolgreichen Übung am 13.12.2012 im Rahmen der 8. International Conference on Geo-information for Disaster Management (Gi4DM) in Enschede und am 31.10.2013 im Rahmen einer Übung am Flughafen Schiphol in den Niederlanden eingesetzt. Hierbei wurden Daten in o.g. Formaten in den Gateway eingespielt und den beteiligten EMS auf Anforderung zur Verfügung gestellt. Ferner konnte eine Anbindung an das Programmsystem DISMA [TÜ15] getestet werden. Hierbei wurden die in Kapitel 4.3 dargestellten Ergebnisse erzielt.

4.2 DISASTER Ontologie

Die Übersetzung bzw. Mediation von spezifischen Informationen der Gefahrenabwehrdomäne wird mit der entwickelten Ontologie EMERGEL (EMERGency ELements) durchgeführt. Im Rahmen eines durchgeführten Expertenworkshops mit Projektpartnern sowie Gästen aus der Gefahrenabwehrdomäne, wurden grundlegende Anforderungen an die Ontologie eruiert. Aus Fragestellungen, wie sie in realen Einsatzlagen entstehen, wurden konkrete Anforderungen extrahiert und soweit möglich mit den verwendeten Datenformaten in eine Tabelle überführt. Grundlegend ist die Ontologie in drei Module gegliedert: Core Modul, transversales Modul und vertikales Modul.

In den Core Modulen werden Schadenslagen als Events klassifiziert. Ein Feuer ist dann als Event klassifiziert, eine Wohnhausbrand, Fabrikbrand oder Waldbrand stellt eine Subklassifizierung dar. Äquivalent dazu wird dieses Konzept der Klassifizierung auf Personen, Einsatzkräfte oder technisches Gerät angewandt. Hier werden dann z.B. Personen als Agent klassifiziert und weitere Subklassen gebildet. Eine mögliche Reihenfolge wäre Person → Feuerwehrmann → Angriffstrupp. Anhand dieser Core Module lassen sich Schadenslagen sowie beteiligte Einheiten repräsentieren.

Die transversalen Module beinhalten den Umgang von EMERGEL mit Raum und Zeit. Großschadensereignisse sind keine statischen Lagen, sondern vielmehr von hoher Dynamik geprägt. So ändert sich bei einem Waldbrand z.B. laufend die Feuerfront (Raum) mit fortschreitendem Ereignis (Zeit).

Die vertikalen Module repräsentieren Konzepte und Vokabular in der Gefahrenabwehrdomäne. Hier sind z.B. die taktischen Zeichen etlicher Länder hinterlegt, Gefahrgutkennzeichnungen, Ländergrenzen oder Orte mit besonderer Bedeutung wie Flughäfen oder Kraftwerke. Mittels der vertikalen Module können z.B. die angeführten taktischen Zeichen „übersetzt“ werden, um sie in der Lagekarte sämtlicher angeschlossener EMS in der jeweils eigenen „Sprache“ verfügbar zu machen [Ru13].

4.3 Beispiel

Anhand eines Beispiels soll das Zusammenspiel zwischen DISASTER Gateway und EMERGEL Ontologie dargestellt werden. Angenommenes Szenario ist hierbei ein Großbrand in einem Moorgebiet zwischen Deutschland und den Niederlanden. Beide Länder entsenden Einheiten in das Schadensgebiet, welche das Feuer an unterschiedlichen Fronten bekämpfen. Gleichzeitig werden in den rückwärtigen Führungseinrichtungen Lagebilder anhand an einer Lagekarte erstellt. Auf beiden Seiten werden die Verläufe der Feuerfront, der eigensetzten Einheiten sowie der Führungsstellen vor Ort mithilfe der jeweiligen taktischen Zeichen dargestellt. Beide Länder verfügen nun über ein Lagebild, dass ihre jeweilige Lage widerspiegelt. Ziel der DISASTER Lösung ist nun, beiden Seiten ein länderübergreifendes Lagebild zu ermöglichen, jedoch in ihrer eigenen „Sprache“, in diesem Fall also ihrer jeweils spezifischen taktischen Zeichen. Die deutsche Seite nutzt für ihre Lagedarstellung einen WFS Server (Web Feature Service), während auf niederländischer Seite ein WMS Server (Web Map Service) genutzt wird. Folgende Prozesse beim Austausch eines taktischen Zeichens werden nun durchgeführt:

- Der Mediator des niederländischen EMS (MediatorNL) sendet eine Anfrage nach der Lagekarte (GML Format).
- Der Mediator des deutschen EMS (MediatorD) erhält die Anfrage und sendet die Daten vom WFS Server.
- Der DISASTER Core transformiert das GML Format in RDF.
- Dieses RDF wird dem MediatorNL zur Verfügung gestellt, dieser generiert mithilfe des DISASTER Core aus dem RDF ein PNG, das Format, in dem die niederländischen Zeichen dargestellt werden.
- Der MediatorNL generiert nun eine WMS Antwort und stellt diese dem niederländischen System zur Verfügung [Ca13].

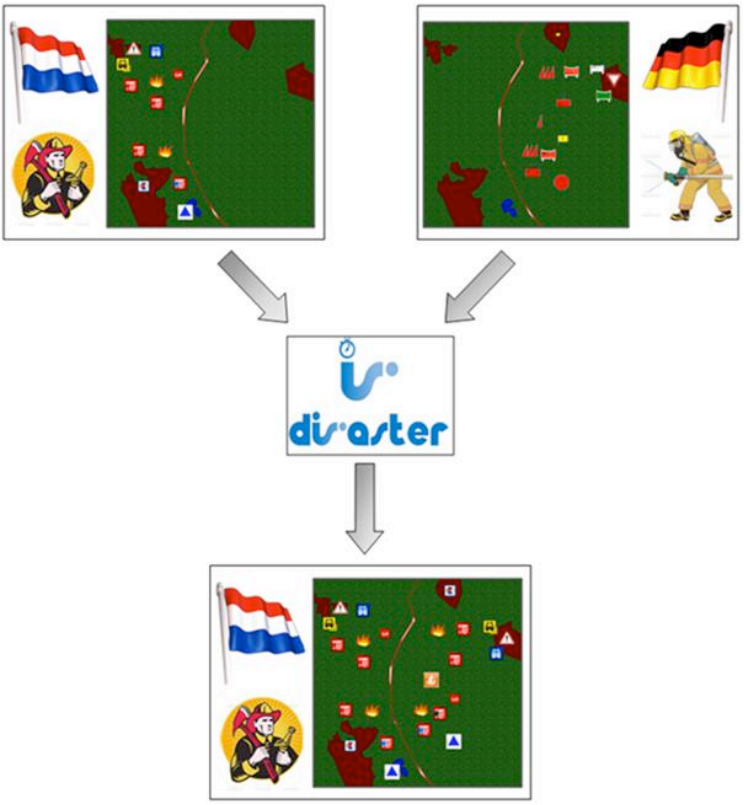


Abb. 1: Lagedarstellung unter Einsatz des DISASTER Gateway [Ca13]

Als Ergebnis der mittels DISASTER Gateway und Ontologie durchgeführten Prozesse steht nun beiden Seiten ein umfassendes, einheitliches Lagebild in ihrer jeweils eigenen „Sprache“ zur Verfügung, wie Abb.1 verdeutlicht.

5 Fazit

Die vom Konsortium des Projekts DISASTER entwickelte Lösung stellt eine Möglichkeit dar, grenzüberschreitenden Informations- und Datenaustauch signifikant zu verbessern. Die Lösung zeichnet sich besonders dadurch aus, dass keine zusätzlichen Systeme beschafft und integriert werden müssen, sondern dass bereits vorhandene und etablierte Systeme untereinander vernetzt werden. Dies birgt mehrere Vorteile: es müssen keine finanziellen Mittel für die Beschaffung neuer Soft- bzw. Hardware

herangezogen werden. Außerdem ist der Schulungsbedarf für die Endanwender gering, da sie weiterhin mit ihren gewohnten Systemen arbeiten. Die auszutauschenden Informationen werden automatisch an die jeweilig benötigten Datenformate angepasst. Außerdem findet eine Übersetzung bzw. Mediation der Daten statt. So sind z.B. taktische Zeichen mittels des Mediationsprozesses direkt nutzbar und vereinfachen eine grenzübergreifende Lagedarstellung erheblich.

Anmerkung:

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 285069.

Literaturverzeichnis

[Ar98] Arbeitsgemeinschaft der Leiter der Berufsfeuerwehren in der Bundesrepublik Deutschland: Qualitaetskriterien für die Bedarfsplanung von Feuerwehren in Städten, 1998.

[Be12] Beuth Verlag GmbH: Informations-/Kommunikationstechnik. DIN SPEC 91287. http://www.beuth.de/de/technische-regel/din-spec-91287/152988042, 10.06.2015.

[Be14] Beuth Verlag GmbH: Medizintechnik/Lebensmittel. DIN EN 1789:2014-12 http://www.beuth.de/de/norm/din-en-1789/222303920, 12.06.2015.

[Ca13] Casado, R.: DISASTER Project. Reference archtitecture & data model approach overview - V1, 2013.

[Em15] Emric+: Projektgruppe. http://emricplus.eu/index.php/de/heim/projektgruppe, 10.06.2015.

[Eu14] European Commission: Humanitarian Aid and Civil Protection. Monitoring Tools. http://ec.europa.eu/echo/node/830, 15.04.2015.

[Eu15] European Commission: Humanitarian Aid and Civil Protection. Emergency Response Coordination Centre (ERCC). http://ec.europa.eu/echo/what/civil-protection/emergency-response-coordination-centre-ercc_en, 14.04.2015.

[Fr10] Fritzen, B.: Die Struktur des Bevölkerungsschutzes in der Bundesrepublik Deutschland. In (Bundesamt für Bevölkerungsschutz und Katastrophenhilfe Hrsg.): Drei Ebenen, ein Ziel: BEVÖLKERUNGSSCHUTZ … gemeinsame Aufgabe von Bund, Ländern und Kommunen, Bonn, 2010; S. 10–15.

[HP02] Holle, P. M.; Pohl-Meuthen, U.: Rettunsdienst im Großschadensfall. In Bundesanstalt für Straßenwesen (Hrsg.): Berichte der Bundesanstalt für Straßenwesen. Mensch und Sicherheit. Wirtschaftsverlag NW, Bremerhaven, 2002.

[Lv11] Luiz, T.; van Lengen, R. H.: Vereinheitlichung von Leitstellenstrukturen und -prozessen als Teil des Qualitätsmanagements im rheinland-pfälzischen Rettungsdienst. In Notfall + Rettungsmedizin, 2011, 14; S. 180–186.

[Ne12] Neuhaus, C. et al.: Crisis management systems in Germany. A status report about the current functions and developments of private and public crisis management systems in Germany. In Rothkrantz, L.; Ristvej, J.; Franco, Z. (Hrsg.): Proceedings of the 9th International ISCRAM Conference, 2012.

[Ru13] Rubiera Azcona, E.: DISASTER Project. EMS Core Ontology - V2, 2013

[St10] Ständige Konferenz für Katastrophenvorsorge und Bevölkerungsschutz: Empfehlungen für Taktische Zeichen im Bevölkerungsschutz, 2010.

[TÜ15] TÜV Rheinland AG: DISMA – Disaster Management. http://www.tuv.com/de/deutschland/gk/anlagen_maschinen/industrieanlagen/disma_disaster_management/disma_disaster_management.html, 10.06.2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Social Media im Krisen- und Katastrophenmanagement - Ein Überblick

Frank Geyer[1], Dr. Volkmar Schau[2] und Prof. Dr. Wilhelm R. Rossak[3]

Abstract: Um die Arbeit von Behörden und Organisationen mit Sicherheitsaufgaben (BOS) im Falle von Krisen und Katastrophen zu vereinfachen, werden in Forschung und Industrie stetig neue IT-Lösungen konzipiert und entwickelt. Ein recht neues Themenfeld in diesem Bereich ist die Nutzung von Social Media (SM)-Angeboten als Informationsquelle und -senke zur Krisenbewältigung. Als Basis für eigene Arbeiten zur Nutzung von SM durch BOS wurde eine Literaturrecherche durchgeführt. Dieses Paper gibt einen Überblick über mögliche Nutzungsfelder und Hürden von SM im Krisen- und Katastrophenmanagement, die aktuelle Situation in der Praxis und den aktuellen Stand der Forschung. Abschließend werden vertiefende Themenfelder identifiziert, deren Bearbeitung in kommenden Schritten erfolgen soll.

Keywords: Notfallmanagement, Social Media, Resilienz, Informationsaustausch, Krisenkommunikationssysteme, Führungskräfteunterstützung

1 Einleitung

Bei großen Schadenslagen, wie Hochwassern oder Massenanfällen von Verletzten, existiert seitens der Bevölkerung und der Medien ein erhöhtes Informationsinteresse. Die Menschen möchten wissen, was geschehen ist und welche Folgen für das eigene Umfeld zu erwarten sind. Auch seitens der Behörden und Organisationen mit Sicherheitsaufgaben (BOS) besteht ein hohes Interesse, Personen in den betroffenen Regionen mit Informationen zu versorgen, um so die Bevölkerung entsprechend vorzubereiten und stärkere Schäden durch Fehlverhalten zu vermeiden. Gerade durch die schnelle Verbreitung von Informationen in neuen Medien besteht zudem der Bedarf, vertrauenswürdige Informationen in den Pool aus mehr oder minder glaubhaften Aussagen einzustreuen.

Zielstellung der Arbeiten des Lehrstuhls für Softwaretechnik der Friedrich-Schiller-Universität Jena in diesem Bereich ist die Entwicklung einer Architektur, mit deren Hilfe der Informationsaustausch zwischen BOS, Medien und der Bevölkerung mit Hilfe von Sozialen Netzwerken entscheidend vereinfacht werden kann. Als Ausgangspunkt für dieses Vorhaben wurde eine Literaturrecherche durchgeführt, um den aktuellen Stand in Theorie und Praxis einschätzen und so Anknüpfungspunkte für die eigene Arbeit finden zu können. Die Ergebnisse der Recherche werden in diesem Paper vorgestellt.

[1] Friedrich-Schiller-Universität Jena, Lehrstuhl für Softwaretechnik, Ernst-Abbe-Platz 2, 07743 Jena, frank.geyer@uni-jena.de
[2] volkmar.schau@uni-jena.de
[3] wilhelm.rossak@uni-jena.de

Die zu untersuchenden Kommunikationskanäle/Medien werden unter dem Begriff Social Media (SM) zusammengefasst. Die Definition dieses Begriffes wird, wie in vielen anderen Publikationen auch, von Andreas M. Kaplan und Michael Haenlein übernommen:

> *„Social Media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content."*[KH10]

Unter diese Definition fallen neben den im allgemeinen Sprachgebrauch unter SM zusammengefassten Plattformen, wie Twitter, Facebook oder Google+ auch Videoplattformen, wie Youtube und Myvideo sowie Angebote wie Blogs, Foren oder Wikis.

2 Nutzungsfelder von Social Media im Krisen- und Katastrophenmanagement

Durch die architektonischen Eigenschaften bietet SM einige Vorteile gegenüber klassischen Medien. Auch im Umfeld des Krisen- und Katastrophenmanagements werden so Nutzungsfelder vor, während und nach eines Ereignisses erkannt.

Ein wichtiger Vorteil ist nach [Fl14] die deutlich höhere Aktualität der verfügbaren Informationen. Dadurch eignen sich neuen Medien besonders gut für das Absetzen von Warnungen an und Hilfesuchen durch die Bevölkerung – gerade wenn klassische Medien, wie Notrufleitungen, zu überlasten drohen [Ka13]. Wegen der hohen Aktualität ist zudem eine schnelle Verbreitung von Informationen zur aktuellen Lage in Krisensituationen möglich.

Ein weiterer Vorteil ist der direkte Kontakt zur Bevölkerung, wodurch eine Informationsgewinnung durch „Menschliche Sensoren" [Ka13] denkbar ist. Diese Art der Anreicherung des Wissensbestandes durch die Bevölkerung wird unter dem Begriff „Social Media Monitoring" zusammengefasst [Ev14]. Ziele hierbei sind neben der Vervollständigung der Lagefeststellung, das Überwachen der Bevölkerungsreaktionen (um Falschmeldungen zu finden und aufklären zu können) sowie das frühzeitige Erkennen von potenziellen Krisen [Gr14]. Dabei wird zwischen aktivem und passivem Monitoring unterschieden: Während das passive Monitoring nur die verfügbaren Informationen der Bevölkerung rezipiert, erfolgt bei aktivem Monitoring eine direkte Kommunikation, etwa durch eigene Statusmeldungen oder durch Antwort auf „Postings" von Einzelpersonen.

Im Zusammenhang mit dem direkten Kontakt zur Bevölkerung liegt ein besonderer Fokus sowohl in der Forschung als auch in der Praxis in der Gewinnung von freiwilligen Helfern durch SM. Vergangene Krisen in aller Welt haben gezeigt, wie schnell Freiwillige sowohl für den virtuellen als auch realen Einsatz durch neue Medien gefunden werden können [RHP13].

Ferner kann SM auch zur Stärkung der Resilienz der Bevölkerung im Vorfeld einer Krise genutzt werden, etwa indem Handlungsempfehlungen zur Vorbereitung auf drohende Gefahren durch BOS veröffentlicht werden [Gr14].

Schließlich ist eine Verwendung von SM-Daten auch im Nachgang an eine Krise im sogenannten „Post Disaster Assessment“ möglich, etwa indem der zeitliche und geografische Verlauf eines räumlich weit verteilten Ereignisses mittels der in Sozialen Netzwerken vorhandenen Daten nachgezeichnet werden kann [Da14].

3 Aktuelle Situation in Deutschland

SM-Angebote gewinnen im privaten Umfeld immer mehr an Bedeutung. So besuchten zum Beispiel 33,36 Millionen Unique Visitors aus Deutschland im September 2013 die Website Facebook[4]. Wegen dieser breiten Nutzung ist SM auch aus Sicht der öffentlichen Hand ein interessanter Kommunikationskanal. Die starke Beschäftigung mit dem Thema, etwa durch das Bundesamt für Bevölkerungsschutz und Katastrophenhilfe (BBK) bestätigt dies.[BB13]

Der Blick auf die internationale Bühne zeigt viele Beispiele, wie stark SM-Angebote auch im Krisenfall genutzt werden (z. B. wurden bis zu 5.500 Tweets pro Sekunde zur Nuklearkatastrophe von Fukushima abgegeben [Me13]) und wie eine gewinnbringende Integration von SM-Anwendungen in die Krisenkommunikation ausgestaltet werden kann (z. B. durch die Integration der Anwendung „Ushahidi“ in das Krisenmanagement während des Erdbebens 2010 in Haiti[5]). Auch in Deutschland gibt es prominente Nutzungsbeispiele: So wird immer wieder der positive Effekt der zum Elbe-Hochwasser 2013 entstandenen Krisenkarte über Google Maps hervorgehoben [Ev14][Mi13]. Weiterhin kann die Helfergenerierung und -verwaltung der Organisation Deutsches Rotes Kreuz (DRK) in Mecklenburg-Vorpommern als praktisches Nutzungsbeispiel genannt werden [Ev14].

So ist auch nicht verwunderlich, dass schon im Januar 2014 über 800 Städte und Kommunen und über 150 Organisationen der Öffentlichen Hand nach eigenen Angaben über Twitter und Facebook aktiv sind [Ev14]. Des Weiteren zeigen mehrere (nicht-repräsentative) Umfragen, dass auch in Deutschland ca. 50 % der Krisenmanager Soziale Netzwerke im Falle einer Katastrophe einsehen, um neue Informationen zu gewinnen und dass auch nicht wenige von diesen gerne in direkte Kommunikation mit der Bevölkerung treten würden [HP13][Fl14].

Trotz dieser durchaus erkennbaren Verbreitung der Nutzung von SM durch die Öffentliche Hand hinkt Deutschland dem internationalen Stand hinterher. So zeigt zum Beispiel [He14a], dass SM vor allem für allgemeine Öffentlichkeitsarbeit (Werbung) oder passives Monitoring (also manueller Suche nach Informationen, ohne in direkte Kommunikation zu treten) genutzt wird. Eine Verwendung von Softwarelösungen, etwa um die Flut an Informationen zu beherrschen, findet selten bis gar nicht statt [He14a][Gr14].

Zusammenfassend lässt sich sagen, dass die öffentliche Hand zwar die Potenziale von SM als Kommunikationsmedium erkennt, jedoch nur einen Bruchteil dieser aktuell tatsächlich nutzbar macht. Gründe hierfür liegen in den speziellen Anforderungen und Hürden, die es noch zu überwinden gilt.

[4] `http://tinyurl.com/ofsvjmy`, zugegriffen am 18.02.2015
[5] `http://tinyurl.com/noyffua`, zugegriffen am 18.02.2015

4 Anforderungen und Hürden

Wenn in einem Bereich neue Systeme eingeführt werden sollen, bestehen naturgemäß spezielle Anforderungen und Hürden, die durch die Technik adressiert bzw. überwunden werden müssen. Vor allem interessant in diesem Bereich sind die wissenschaftlichen Arbeiten von [Hu14] und [DPA14] die (nicht repräsentative) Befragungen und Experteninterviews durchgeführt haben, um grundlegende Anforderungen zu erheben. Das Thema wird aber auch an anderen Stellen (etwa [HKP14] oder [Le14]) adressiert. Zusammenfassend wurden folgende Hauptanforderungen in der Literatur identifiziert:

- Neue Formen von SM entstehen aktuell sehr schnell, sodass entsprechende Software leicht auf neue Kanäle adaptierbar sein muss.
- Da verschiedene SM-Kanäle verschiedene Vor- und Nachteile mit sich bringen, müssen sowohl mehrere Kanäle gleichzeitig bedient werden, als auch jeweils an den Kanal angepasste Kommunikationsstrategien definiert werden.
- Im Krisenfall kann es zu Ausfällen der Infrastruktur kommen, sodass entsprechende Software offlinefähig sein muss.
- Durch die zu bewältigende Informationsflut werden in Anwendungen dringend (semi-) automatische Methoden zur Sortierung und Filterung benötigt. Der dahinterliegende Prozess der Aggregation und Komposition muss jedoch transparent gehalten werden. Außerdem müssen Darstellungen gefunden werden, die gleichzeitig möglichst viele Informationen zeigen, aber auch dem Nutzer einen einfachen Überblick ermöglichen.
- Innerhalb des Monitorings wünschen sich die Endanwender meist eine Trennung der Informationen nach Kanal. Eine Aggregation über mehrere Kanäle soll nur in Spezialfällen (etwa der Suche nach allem verfügbaren Bildmaterial) erfolgen.
- Ein großer Nutzen wird in der Auswertung von ortsbezogenen Daten gesehen, durch die automatisiert Kartendarstellungen erzeugt werden können. Grundlage hierfür ist, dass viele mobile Anwendungen zur Nutzung von SM die Übersendung von Geodaten unterstützen.
- Zur Anreicherung und Verifizierung von Informationen muss eine Verbindung zu anderen Datenquellen (z.B. Wetterdaten) und eine Integration in die bestehende Infrastruktur möglich sein.
- Die Erstellung automatischer Reports nach Vorlage klassischer Pressespiegel ist ein häufig geäußerter Wunsch.
- Wichtig ist ebenfalls eine möglichst vollständige Dokumentation aller SM-Aktivitäten der BOS, um so nicht nur eine Rechtssicherheit sondern auch eine nachgehende Auswertbarkeit zu erreichen.
- Für die Ausgabe von Informationen über Soziale Netzwerke (Informationssender) werden entsprechende Vorlagen für wiederkehrende Nachrichten gewünscht (z.B. Meldung des aktuellen Hochwasserstandes).

Zur Erfüllung dieser und ähnlicher Anforderungen müssen eine Reihe an Hürden überwunden werden, um eine umfassende Nutzung von SM im Krisen- und Katastrophenmanagement zu erreichen. Viele Autoren haben sich in den letzten Jahren mit diesem Thema befasst. Exemplarisch seinen die Arbeiten von [He14b], [HP13], [Ho14] und [Ri13] genannt. In Summe lassen sich in Anlehnung an [Be13] vier Hauptkategorien definieren, um die bestehenden Hürden zu klassifizieren.

4.1 Technologische Hürden

Allein die Flut an vorhandenen Informationen stellt eine große Hürde für deren Nutzung dar. Wenn zu einem Ereignis wie dem Hurricane Sandy im November 2012 2 Mio. Tweets veröffentlicht werden [Me13] ist das eine Informationsmenge, die händisch nur sehr schwer zu sichten und zu bewerten ist. Außerdem stellt die Vertrauenswürdigkeit ein großes Hindernis dar. Da der Bürger i.d.R. keine journalistische Ausbildung aufweisen kann, werden Informationen immer wieder ohne ausreichende Prüfung als wahr angenommen und verbreitet.

Auch die bereits unter den Anforderungen genannten Punkte der Offlinefähigkeit, Adaptierbarkeit an neue Medien und Integration in die bestehende IT-Landschaft gehören zu den zu nennenden technologischen Hürden.

4.2 Organisationale und Ökonomische Hürden

Unter den wesentlichsten Hemmnissen für eine schnelle Einführung von SM wird immer wieder fehlendes Geld und fehlendes Personal seitens der BOS genannt. Aktuelle Haushaltspläne und -sperren stellen nach [PB14] sogar das größte Hindernis dar. Einhergehend mit diesem Personalmangel sind auch Veränderungen an den etablierten Strukturen notwendig: Nicht jeder Beitrag, der auf einer SM-Plattform veröffentlicht wird, kann den klassischen Freigabeprozess durchlaufen, weil dadurch eine schnelle Reaktion nicht möglich ist. Es wird also ein neues Rollenkonzept und ein angemessener Freigabeprozess benötigt. Ganz grundsätzlich fehlt es aktuell an genügend klaren Regeln, best practices und Standardverfahren, wie die Nutzung von SM durch eine BOS möglichst effektiv und zielgerichtet erfolgen kann.

4.3 Soziale Hürden

Auch wenn Soziale Netzwerke heute über eine beachtliche Reichweite verfügen, können nicht alle Schichten der Bevölkerung gleichmäßig erreicht werden. Es besteht die Gefahr der Entstehung von blinden Flecken, also bestimmten Gebieten oder Bevölkerungsschichten, die nicht erreicht, bzw. deren Probleme nicht gehört werden. Außerdem befürchten viele Behörden einen Kontrollverlust über die Aktivitäten in Sozialen Netzwerken einhergehend mit der Gefahr von Selbstjustiz, wie man sie bei den Aufruhren in London [Wa14]

und Vancouver [Ri13] erleben konnte. In diesem Zusammenhang steht auch die fehlende Sensibilität in der Bevölkerung bezüglich der Konsequenzen, wenn persönliche Daten in Sozialen Netzwerken verbreitet werden.

Außerdem ist unter dem Bereich der sozialen Aspekte die durch die BOS wahrgenommene geringe Hilfswilligkeit zu nennen: Die Experten aus der Praxis sprechen von einer stark gestiegenen Anzahl an Notrufen wegen Lappalien, gerade im städtischen Raum. Demgegenüber stehen jedoch die positiven Erfahrungen mit virtueller und realer Hilfe zum Beispiel beim Elbe-Hochwasser 2013.

4.4 Rechtliche Hürden

Der Katastrophenschutz ist nach §§30,70 des Grundgesetzes Landesrecht, sodass neben den relevanten Bundesgesetzen auch die jeweiligen Landesgesetzgebungen beachtet werden müssen[6][7]. Die Bandbreite an relevanten Gesetzen und Verordnungen, die bei einer Nutzung von SM im Krisenfall zu beachten sind, reicht vom jeweiligen Rettungsdienstgesetz und der geltenden Feuerwehrdienstvorschrift über datenschutzrechtliche und medienrechtliche Aspekte (Telemediengesetz) bis hin zu vertrags- und vergaberechtlichen Fragestellungen. Die schiere Menge an berührten Gesetzen schreckt viele BOS bereits von der Verwendung ab. Hinzu kommt, dass das Grundgesetz kein „Not kennt kein Gebot“ kennt [Ev14] und das viele rechtliche Fragestellungen aktuell noch nicht abschließend geklärt wurden (etwa die Haftung für geteilte Inhalte [Di14]). Erschwerend kommt weiterhin hinzu, dass viele der Anbieter von SM keine deutschen Unternehmen sind, sodass die jeweils nationale und internationale Rechtsprechung beachtet werden muss.

5 Aktueller Forschungsprojekte und -gebiete

National und International wird SM im Bereich der zivilen Sicherheit stark erforscht. Im folgenden Abschnitt sollen einige interessante Forschungsprojekte und deren Zielstellungen genannt werden. Weiterhin soll ein Einblick in die Forschungen zu Facebook und Twitter im Speziellen und zu den Themen der Integration freiwilliger Helfer und zu Richtlinien für die Nutzung von SM durch BOS im allgemeinen gegeben werden, weil diese Bereiche zu den internationalen Forschungsschwerpunkten gehören und/oder besondere Bedeutung mit Blick auf die Situation in Deutschland haben.

5.1 Aktuelle Forschungsprojekte zu SM

An dieser Stelle seien einige interessante Forschungsprojekte genannt, deren Ergebnisse für die eigenen Arbeiten von Relevanz sein können. Die Liste ist in alphabetischer Reihenfolge und ist nicht als vollständig zu verstehen, sondern dient lediglich für einen Einblick in die aktuelle Förderlandschaft in diesem Bereich.

[6] `http://dejure.org/gesetze/GG/30.html`, zugegriffen am 18.02.2015
[7] `http://dejure.org/gesetze/GG/70.html`, zugegriffen am 18.02.2015

- COSMIC[8]: Das Projekt „Contribution of Social Media in Crisis management“ wird durch die EU mit rund 1,2 Millionen Euro gefördert und untersucht in einem Zeitraum von April 2013 bis März 2015 unter anderem, welche Leitlinien für die Nutzung Sozialer Netzwerke sowohl für Bevölkerung als auch für private und öffentliche Organisationen formuliert werden können. Zielstellung ist die Identifikation effektiver Möglichkeiten zur Nutzung Sozialer Netzwerke durch die Bevölkerung und die öffentliche Hand im Falle einer Krise. Das Konsortium besteht aus sieben Partnern aus sechs Nationen und umfasst neben universitären Einrichtungen auch IKT-Unternehmen und Endanwender.
- EMERGENT[9]: Das Projekt „Emergency Management in Social Media Generation“ wird durch die EU mit rund 2,7 Millionen Euro gefördert und untersucht in einem Zeitraum von April 2014 bis März 2017 unter anderem die aktuelle Nutzung von SM im Krisenfall, um organisatorische und technische Leitlinien für die Zukunft ableiten zu können. Bereits im Vorgänger-Projekt des Bundesministerium für Bildung und Forschung (BMBF) mit dem Namen „infostrom“ ist dabei mit `www.sicherheitsarena.org` eine Plattform entstanden, die zur Verbesserung der Informationserfassung, -aufbereitung und -weitergabe innerhalb und zwischen BOS, den Medien, der Bevölkerung und sonstigen Beteiligten dient [in13].
- INSIGHT[10]: Das Projekt „Intelligent Synthesis and Real-time Response using Massive Streaming of Heterogeneous Data“ wird durch die EU mit rund 2,8 Millionen Euro gefördert und untersucht in einem Zeitraum von September 2012 bis August 2015 unter anderem, wie durch Big-Data-Analysen (auch mit den in Sozialen Netzwerken vorhandenen Informationen) eine Früherkennung von Krisen möglich ist.
- ISAR+[11]: Das Projekt „Online and Mobile Communications for Crisis Response and Search and Rescue“ wird durch die EU mit rund 3,8 Millionen Euro gefördert und untersucht in einem Zeitraum von Januar 2013 bis Juni 2015 unter anderem, wie die Mitwirkung der Bevölkerung im Krisenmanagement verbessert werden kann. Zielstellung in diesem Bereich ist neben Leitlinien für BOS und die Bevölkerung auch die Schaffung einer entsprechenden Plattform.
- RESCUER[12]: Das in Kooperation mit Brasilien durchgeführte Projekt „Reliable and Smart Crowdsourcing Solution for Emergency and Crisis Management“ wird durch die EU mit rund 1,3 Millionen Euro gefördert und untersucht in einem Zeitraum von Oktober 2013 bis März 2016, wie durch Crowd-Sourcing eine schnellere Hilfe bei Problemen während Massenveranstaltungen erreicht werden kann. Dazu sollen zum Beispiel automatische Videoanalysen des auch in Sozialen Medien vorhandenen Materials und das Tracking der vor Ort vorhandenen Mobiltelefone zum Einsatz kommen. Die praktische Erprobung des Systems findet zur Fußball-Weltmeisterschaft 2014 und zu den olympischen Spielen 2016 in Brasilien statt.

[8] `http://www.cosmic-project.eu`, zugegriffen am 18.02.2015
[9] `http://www.fp7-emergent.eu`, zugegriffen am 18.02.2015
[10] `www.insight-ict.eu`, zugegriffen am 18.02.2015
[11] `http://www.isar.i112.eu`, zugegriffen am 18.02.2015
[12] `http://www.rescuer-project.org`, zugegriffen am 18.02.2015

- SIKOMM[13]: Das Projekt „Sicherheitskommunikation: Strategien zur kommunikativen Unterstützung von Schutz- und Rettungsmaßnahmen“ wurde durch das BMBF mit rund 600.000 Euro gefördert und untersuchte in einem Zeitraum von Februar 2009 bis Juli 2012 die Kommunikationsprozesse zwischen allen, im Krisenmanagement relevanten Akteuren. Auch hier war es Ziel, verallgemeinerbare kommunikative und mediale Strategien zu entwickeln, die den Informationsaustausch zwischen allen Beteiligten erleichtern.
- VOTEKK[14]: Das Projekt „Vorbereitung auf Terroranschläge, Krisen und Katastrophen“ wurde durch das BMBF mit rund 3 Mio. Euro gefördert und erarbeitete in einem Zeitraum von Juni 2009 bis Mai 2012 eine eLearning-Plattform für Vorbereitung auf Katastrophen. Auch die Bevölkerung soll durch diese Plattform besser auf Krisen vorbereitet werden.

5.2 Twitter als Krisenmedium

Weil jeder Tweet vollständig öffentlich einsehbar und weil Twitter für Entwicklungen durch Dritte sehr gut geeignet ist (z. B. bietet Twitter eine API an[15]) fokussieren viele softwarenahe Forschungen auf dieses Medium. Die Bandbreite der Forschungen reicht von Untersuchungen zur Nutzung von Twitter bei vergangenen Ereignissen (z. B.: [Te12], [Ya13], [WF14]) bis hin zu real einsetzbaren Anwendungen, die Twitter als Kommunikationsmedium nutzen.

Bei keinem anderen Sozialen Netzwerk wird dabei die schiere Informationsflut, die zu bewältigen ist, so deutlich. Zur Überwindung dieser Informationsmassen beschäftigt sich die Forschung zum einen mit der automatischen Analyse und Bewertung von Tweets (um so Informationen automatisiert zu filtern und zu gewichten) als auch mit der grafischen Aufbereitung der Informationen (zum Beispiel in Form von Karten, Tagclouds oder Zeitcharts). Entsprechende Arbeiten hierzu sind unter anderem [HP13], [WKS13], [SKW13], [SBP14] und [Pu14]. Neben den automatischen Verfahren zur Filterung und Bewertung sowie zur Verifikation getweeteter Informationen wird auch der Einsatz der Crowd (also vieler freiwilliger Helfer) diskutiert und auch praktisch genutzt: Ein Beispiel hierfür ist die Vorgabe zu verwendender Hashtags durch große NGO[16] und die damit verbundene Aufbereitung vorhandener Tweets durch Freiwillige [HP13]. Des Weiteren werden die Möglichkeiten von Twitter auch in der Nachbereitung von Ereignissen und für die Aufklärung der Bevölkerung untersucht [Da14].

Am interessantesten sind jedoch die vorhandenen Anwendungen die auf Twitter basieren: Twitcident etwa, das Tweets während eines Ereignisses automatisiert filtert und grafisch aufbereitet, sodass Krisenstäbe einen Überblick über aktuelle Themen im Sozialen Netzwerk erhalten [Te12]. Ushahidi, dass genutzt werden kann, um Hilfegesuche und Hilfsan-

[13] `http://www.uni-siegen.de/ifm/projekte/sikomm/`, zugegriffen am 18.02.2015
[14] `http://www.votekk.de`, zugegriffen am 18.02.2015
[15] `https://dev.twitter.com/rest/`, zugegriffen am 18.02.2015
[16] Nichtregierungsorganisationen wie z. B. der Deutsche Caritasverband e.V.

gebote direkt innerhalb der Bevölkerung oder auch mit NGO's zu teilen [17]. GDACSmobile, dass ein bestehendes globales Katastrophenwarnsystem durch eine SM-Komponente erweitert [LHG13]. Einige weitere Anwendungen könnten genannt werden.

5.3 Facebook als Krisenmedium

In Deutschland ist Facebook das am weitesten verbreitetste Soziale Netzwerk. Daher engagieren sich die meisten Nutzer der öffentlichen Hand auch in diesem Netzwerk: Verschiedene Studien sprechen von bis zu 95 % an Nutzern von Facebook unter den Nutzern von SM der öffentlichen Hand [He14a][Ho14]. Auch Facebook bietet eine Programmierschnittstelle an, über die Drittsysteme zugreifen können[18]. Dennoch ist Facebook in der aktuellen Forschung weniger häufig Gegenstand. Die verfügbaren Arbeiten untersuchen meist die Nutzung vergangener Ereignisse (etwa [RnGP14]) und leiten aus diesen Hinweise für zukünftige Nutzungen ab [TW13]. Die Ursachen für diese geringere Auseinandersetzung mit Facebook als Medium für das Krisen- und Katastrophenmanagement müssen noch ergründet werden. Mögliche Erklärungen könnten in der nicht vollständigen Einsehbarkeit aller Informationen oder in der Sperrung der Nutzung von Facebook in einigen Krisenregionen durch die dortige Regierung liegen. Ein weiterer Grund könnten die Datenschutzrichtlinien von Facebook sein, die das Recht an quasi jeder verfügbaren Information an Facebook übertragen.

5.4 Integration freiwilliger Helfer

Das auch von den deutschen Einsatzkräften erkannte Potenzial der Gewinnung zusätzlicher freiwilliger Helfer durch SM wird in der Forschung bereits untersucht. Vor allem in den Phasen der Einsatzabarbeitung, in denen noch nicht genügend ausgebildetes Personal vor Ort ist, können durch freiwillige Helfer wichtige Arbeiten übernommen werden. Die Untersuchungen unterscheiden dabei meist zwischen virtuellen und realen Helfern, die entweder koordinierend und aufbereitend am PC zum Beispiel die Informationsflut bearbeiten oder aktiv am Schadensort Aufgaben übernehmen [HP13] [RHP13]. Für beide Fälle werden Systeme zur besseren Koordinierung und damit Nutzung des freiwilligen Hilfspotenzials untersucht.

5.5 Richtlinien für die Nutzung von SM

Neben den bereits genannten Forschungsprojekten, die sich unter Anderem mit der Erstellung von Richtlinien für die Nutzung von SM durch die öffentliche Hand beschäftigen, wird dieser Bereich in [Ka13], [Ev14], [Gr14] und [Ho14] thematisiert. Die Erkenntnisse dort lassen sich wie folgt zusammenfassen:

[17] `www.ushahidi.com`, zugegriffen am 18.02.2015

[18] `https://developers.facebook.com/docs/apis-and-sdks/`, zugegriffen am 18.02.2015

- In der Regel erfolgt der erste Schritt von BOS in Sozialen Netzwerken „blind“, d.h. ohne eine vorab (niedergeschriebene) SM-Strategie. Im Bestfall trifft man auf Guidelines, die an Netiquette-Vorgaben erinnern. Ein gutes Beispiel hierfür sind die „Social Media Guidelines“ des THW[19].
- Wichtig ist in jedem Fall die Informations- und Interpretationshoheit zu behalten. Dazu muss sich die BOS frühzeitig an Diskussionen in SM beteiligen und schon im Vorfeld zu einer Katastrophe ein Netzwerk und damit Vertrauen aufbauen.
- BOS sollten mehrere SM-Kanäle bedienen und regelmäßig und ereignisunabhängig Informationen (Veranstaltungshinweise, Personalsuche, Nachwuchswerbung, Neuigkeiten zur Organisation, Maßnahmen zur Prävention, Verhalten im Krisenfall) veröffentlichen, um so Vertrauen in das Medium und das BOS-Profil zu schaffen.
- Die Auswahl der Medien sollte an Hand der Nutzerzahlen in Deutschland, des Funktionsumfangs, des Zwecks der Plattform und der Nutzungskosten erfolgen. Die getroffene Auswahl ist regelmäßig (jährlich) zu prüfen.
- Die Inhalte verschiedener Kanäle (auch Radio, Fernsehen u.Ä.) müssen synchronisiert werden, sodass keine Widersprüche (und daraus Vertrauensverlust) entstehen.
- Nutzerbeiträge sollten auf dem eigenen Profil zugelassen werden, müssen aber regelmäßig gelesen, beantwortet, moderiert oder im Extremfall gelöscht werden.

6 Weiterführende Arbeiten

Aus den zusammengetragenen Informationen ergeben sich einige Themenfelder, die für die eigenen Forschungsarbeiten vertieft untersucht werden können. Der kommende Arbeitsschritt beinhaltet daher zunächst eine spezialisierte Recherche in diesem Bereich, um auch hier den aktuellen Stand der Technik und der Forschung genauer zu untersetzen und so den aktuellen Forschungsbedarf weiter eingrenzen zu können. Dies sind vor allem:

- Systeme zur Integration freiwilliger Kräfte auf operativer und strategischer Ebene.
- Systeme zur Allokation privater Ressourcen.
- Bewertung verschiedener SM-Plattformen als Warnkonzepte nach [He14b].
- Ermittlung der Potenziale von Facebook als Krisenmedium.
- Vertiefende Aufarbeitung der rechtlichen Rahmenbedingungen in Deutschland.
- Aufbereitung der Möglichkeiten der Programmierschnittstellen der Plattformen.

[19] `http://tinyurl.com/n69sj9o`, zugegriffen am 18.02.2015

Literaturverzeichnis

[BB13] BBK: Social Media - näher am Puls der Katastrophe. Jahresbericht 2013, S. 38–39, 2013.

[Be13] Beneito-Montagut, Roser; Anson, Susan; Shaw, Duncan; Brewster, Christopher: Governmental Social Media use for Emergency Communication. In: 10. ISCRAM. 2013.

[Da14] Dashti, Shideh; Palen, Leysia; Heris, Mehdi P.; Anderson, Kenneth M.; Jennings, T.; Anderson, Scott: Supporting Disaster Reconnaissance with Social Media Data: A Design-Oriented Case Study of the 2013 Colorado Floods. In: 11. ISCRAM. 2014.

[Di14] Dirks, Stephan: Rechtliche Aspekte des Einsatzes von Social Media im Katastrophenmanagement. In: Social Media im Krisen- und Katastrophenmanagement. GITO mbH Verlag, S. 53–62, 2014.

[DPA14] Denis, Lise Ann St.; Palen, Leysia; Anderson, Kenneth M.: Mastering Social Media: An Analysis of Jefferson County's Communications during the 2013 Colorado Floods. In: 11. ISCRAM. 2014.

[Ev14] Evertz, Katja; Lüge, Timo; Zisgen, Julia; Kern, Julia; Voßschmidt, Stefan; Steiger, Dr. Saskia; Schiller, Prof. Dr.-Ing. Jochen; Gerhold, Dr. Lars: Social Media. BBK Magazin Bevölkerungsschutz 3/2014, 2014.

[Fl14] Flizikowski, Adam; Holubowicz, Witold; Stachowicz, Anna; Hokkanen, Laura; Kurki, Taina; Paeivinen, Niina; Delavallade, Thomas: Social Media in Crisis Management – the iSAR+ Project Survey. In: 11. ISCRAM. 2014.

[Gr14] Gronau, Norbert; Heine, Moreen; Thim, Christof: Eckpunktepapier - Social Media im Krisenmanagement, 2014.

[He14a] Heine, Dr. Moreen: Social Media im Krisen- und Katastrophenmanagement - Stand und Verbreitung. In: Social Media im Krisen- und Katastrophenmanagement. GITO mbH Verlag, 2014.

[He14b] Heine, Dr. Moreen: Soziale Medien in der zivilen Sicherheit - Integration in Strukturen und Prozesse. In: Social Media im Krisen- und Katastrophenmanagement. GITO mbH Verlag, S. 63–70, 2014.

[HKP14] Hiltz, Starr Roxanne; Kushma, Jane; Plotnick, Linda: Use of Social Media by U.S. Public Sector Emergency Managers: Barriers and Wish Lists. In: 11. ISCRAM. 2014.

[Ho14] Hofmann, Sara: 'Just Because we can' - Governments rationale for using social media. In: 11. ISCRAM. 2014.

[HP13] Hiltz, Starr Roxanne; Plotnick, Linda: Dealing with Information Overload When Using Social Media for Emergency Management: Emerging Solutions. In: 10. ISCRAM. 2013.

[Hu14] Hughes, Amanda Lee: Participatory Design for the Social Media Needs of Emergency Public Information Officers. In: 11. ISCRAM. 2014.

[in13] infostrom-Abschlussbericht, http://tinyurl.com/owwa9et, zugegriffen am 18.02.2015, 2013.

[Ka13] Karsten, Andreas: Nutzung von Social Media zur Entscheidungsunterstützung. BBK Magazin Bevölkerungsschutz 2/2013, S. 36–38, 2013.

[KH10] Kaplan, Andreas M.; Haenlein, Michael: Users of the world, unite! The challenges and opportunities of Social Media. In: Business Horizons Nr. 53. 2010.

[Le14] Leitlinien des Deutschen Städtetages zur Kommunikation der Städte im Bereich Social Media, http://www.staedtetag.de/presse/beschluesse/069443/, zugegriffen am 18.02.2015, 2014.

[LHG13] Link, Daniel; Hellingrath, Bernd; Groeve, Tom De: Twitter Integration and content moderation in GDACSmobile. In: 10. ISCRAM. 2013.

[Me13] Meier, Patrick Philippe; Castillo, Carlos; Imran, Muhammad; Elbassuoni, Shady Mamoon; Diaz, Fernando: Extracting Information Nuggets from Disaster-Related Messages in Social Media. In: 10. ISCRAM. 2013.

[Mi13] Mildner, Sven: Bürgerbeteiligung beim Hochwasserkampf - Chancen und Risiken einer kollaborativen Internetplattform zur Koordination der Gefahrenabwehr. In: Gemeinschaften in Neuen Medien 2013. 2013.

[PB14] Proske, Niels; Baban, Dr. Constance P.: Soziale Medien im Krisenmanagement - Zusammenfassung und Ausblick. In: Social Media im Krisen- und Katastrophenmanagement. GITO mbH Verlag, S. 71–77, 2014.

[Pu14] Purohit, Hemant; Bhatt, Shreyansh; Hampton, Andrew; Shalin, Valerie; Sheth, Amit; Flach, John: With Whom to Coordinate, Why and How in Ad-hoc Social Media Communities during Crisis Response. In: 11. ISCRAM. 2014.

[RHP13] Reuter, Christian; Heger, Oliver; Pipek, Volkmar: Combining Real and Virtual Volunteers through Social Media. In: 10. ISCRAM. 2013.

[Ri13] Rizza, Caroline; Ângela Guimarães, Pereira; Chiaramello, Michel; Curvelo, Paula: Do-It-Yourself Justice: Considerations of Social Media Use in a Crisis Situation: The Case of the 2011 Vancouver Riots. In: 10. ISCRAM. 2013.

[RnGP14] Rizza, Caroline; Ângela Guimarães Pereira: Building a resilient community through social network: ethical considerations about the 2011 Genoa floods. In: 11. ISCRAM. 2014.

[SBP14] Shalunts, Gayane; Backfried, Gerhard; Prinz, Katja: Sentiment Analysis of German Social Media Data for Natural. In: 11. ISCRAM. 2014.

[SKW13] Schaust, Sven; Kaisser, Michael; Walther, Maximilian: Avalanche: Prepare, Manage, and Understand Crisis Situations using Social Media Analytics. In: 10. ISCRAM. 2013.

[Te12] Terpstra, Teun; de Vries, A.; Stronkman, R.; Paradies, G.L.: Towards a realtime Twitter analysis during crises for operational crisis management. In: 9. ISCRAM. 2012.

[TW13] Tyshchuk, Yulia; Wallace, William: The Use of Social Media by Local Government in Response to an Extreme Event: Del Norte County, CA and the 2011 Japan Tsunami. In: 10. ISCRAM. 2013.

[Wa14] Watson, Hayley; Baruh, Lemi; Finn, Rachel L.; Scifo, Salvatore: Citizen (in)security?: social media, citizen journalism and crisis response. In: 11. ISCRAM. 2014.

[WF14] Watson, Hayley; Finn, Rachel L.: Social media and the 2013 UK heat wave: opportunities and challenges for future events. In: 11. ISCRAM. 2014.

[WKS13] Walther, Maximilian; Kaisser, Michael; Schaust, Sven: Social Media-Based Event Detection for Crisis Management in the Al Za'atari Refugee Camp. In: 10. ISCRAM. 2013.

[Ya13] Yang, Seungwon; Chung, Haeyong; Lin, Xiao; Lee, Sunshin; Chen, Liangzhe; Wood, Andrew; Kavanaugh, Andrea L; Sheetz, Steven D.; Shoemaker, Donald J.; Fox, Edward A.: PhaseVis: What, When, Where, and Who in Visualizing the Four Phases of Emergency Management Through the Lens of Social Media. In: 10. ISCRAM. 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Big Data, Smart Data and Semantic Technologies (BDSDST 2015)

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Big Data, Smart Data and Semantic Technologies (BDSDST 2015)

Natalja Kleiner[1], Suad Sejdovic[2], Stefan Zander[3], Thomas Setzer[4], Rudi Studer[5], Stefan Jähnichen[6]

Abstract: Environmental responsibility, sustainability and energy efficiency are becoming increasingly important in recent times in almost every area of our daily life. At the same time, the growing technical possibilities to gather and aggregate multi-modal data from sensors, mobile devices, social media, log files, cameras, microphones etc. provide useful information about the state and the environment of the data sources. These gathered large data sets today are known as Big Data, which, when processed properly, might carry huge amounts of useful information, which was not accessible beforehand and allow for better-founded, more robust predictions and better decision-making amongst others regarding environmental and energy aspects. Besides new analytic approaches, novel information technologies such as semantic technologies are necessary in order to exploit the full potential of the gathered data. Semantic technologies facilitate the interrelation and integration of data, content, and processes in meaningful ways, which is very important when dealing with high amounts of heterogeneous data.

That is why using Big and Smart Data as well as methods and tools based on semantic technologies can provide more transparency, enable precise and well-founded decisions and improve planning processes, which will result in more sustainable and efficient processes and systems.

This workshop brings together researchers and practitioners working on Big Data and semantic technologies in various fields. We accepted six submissions for presentation, which identify the potentials of Big Data and semantic technolgies in the energy and logistics domain as well as the challenges posed by Big Data on knowledge discovery, task description and data search.

Program Committee: Andreas Abecker (disy GmbH, Karlsruhe, Germany), Boris Amberg (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Matthias Bender (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Martin Birkmeier (FIR e.V. an der RWTH Aachen, Aachen, Germany), Felix Brandt (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Simone Braun (CAS Software AG, Karlsruhe, Germany), Peter Haase (metaphacts GmbH, Walldorf, Germany), Andreas Harth (Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany), Iris Heckmann (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Mark Hefke (CAS Software AG, Karlsruhe, Germany), Marcel Huschebeck (PTV Group, Karlsruhe, Germany), Rico Knapper (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Alexander Lenk (FZI Forschungszentrum Informatik,

[1] FZI Forschungszentrum Informatik, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, natalja.kleiner@fzi.de
[2] FZI Forschungszentrum Informatik, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, sejdovic@fzi.de
[3] FZI Forschungszentrum Informatik, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, zander@fzi.de
[4] Karlsruher Institut für Technologie (KIT), Corporate Services and Systems, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, thomas.setzer@kit.edu
[5] Karlsruher Institut für Technologie (KIT), Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB), Englerstr. 11, 76131 Karlsruhe, studer@kit.edu
[6] FZI Forschungszentrum Informatik, Aussenstelle Berlin, Friedrichstr. 60, 10117 Berlin, stefan.jaehnichen@fzi.de

Branch Office Berlin, Germany), Marco Lewandowski (LogDynamics Lab, Bremen, Germany), Peter C. Lockemann (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Anne Meyer (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Jan Necil (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Stefan Nickel (Karlsruher Institut für Technologie, Karlsruhe, Germany), Jens Nimis (HS Karlsruhe, Germany), Christoph Pinkel (fluidOps, Walldorf, Germany), Achim Rettinger (Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany), Bernd Schönebeck (cognesys gmbh, Aachen, Germany), Alexander Schuller (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Nenad Stojanovic (Nissatech, Nis, Serbia), Ljiljana Stojanovic (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Thomas Usländer (Fraunhofer IOSB, Karlsruhe, Germany), Annika Vernbro (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Walter Waterfeld (Software AG, Darmstadt, Germany), Christof Weinhardt (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Jacqueline Wirnitzer (FZI Forschungszentrum Informatik, Karlsruhe, Germany), Valentin Zacharias (Daimler TSS, Ulm, Germany), Stefan Zander (FZI Forschungszentrum Informatik, Karlsruhe, Germany)

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

From Smart Meters to Smart Products: Reviewing Big Data driven Product Innovation in the European Electricity Retail Market

Nicolai Krüger[1], Frank Teuteberg[2]

Abstract: This paper tries to provide a perspective on energy informatics that goes beyond solely technical research. Combining the viewpoints of energy informatics and information systems in a review of top-ranked literature, we aim to start the discussion how to bridge the gap between a technology-driven focus on big data possibilities (such as smart grid, smart metering, etc.) and a business model-driven perspective (innovation, price-reduction, CRM, etc.). With this paper we further underline the necessity of an interdisciplinary research approach and try to investigate a value chain-oriented implementation of big data initiatives. Especially in the European electricity branch, big data – in terms of smart metering – seems in some cases to be implemented in order to fulfil governmental regulations. CRM as well as corporations' change management has been neglected. By conducting a structured literature review, the status quo of smart meter related big data (initiatives) in energy informatics and information systems will be discovered, and a research agenda for further research in this area will be provided.

Keywords: Big data, change management, big data implementation, smart metering, smart grid, energy informatics, data-driven business, data-driven innovation.

1 Introduction and broader research scope

1.1 Linking big data, change, product innovation and smart metering

Initiated by the European Union's directive on energy efficiency [Eu06], energy providers were obligated to offer demand-based price packages for electricity to consumers, which gave big data the potential to reach every single European household: Typically big data is defined by the 4 V's, volume, velocity, variety and veracity [Go14][MB12]. Measured with smart meter devices (veracity), any electricity consumption will cause data (volume) that needs to be transmitted from the end-user to the energy provider (velocity) [SK12][MRT12]. Matching the smart meter data with additional information resources of the consumer, e.g. weather conditions, would offer energy providers a huge variety of data to create smart and consumption related pricing packages [REG10]. But while the EU's regulations have been fulfilled by launching

[1] University of Osnabrueck, Accounting and Information Systems, Katharinenstr. 1, 49069 Osnabrueck, Germany, an@nicolaikrueger.de
[2] University of Osnabrueck, Accounting and Information Systems, Katharinenstr. 1, 49069 Osnabrueck, Germany, frank.teuteberg@uni-osnabrueck.de

smart grids, smart meters, and a variety of big data technologies to offer individual price ranges for consumers, the real potential of data-driven business models did not become the core focus for electricity providers, as a T-Systems survey among 250 managers in that branch discovered in February 2013: 47% of the interviewees do not see the value within data, even if they see an annual data-growth-rate of 25% [TS13].

Abstracting the given example, the question of how big data should be implemented has become more relevant today, compared to the questions asked in the past regarding whether or not large-scale and high frequency data existed and which mathematical patterns and predictions could be engineered with it. Therefore this paper focuses on reviewing the status quo in energy informatics (EI) and information systems (IS) regarding the combination of three elements: big data, change/product innovation management and smart metering. Smart metering, which is defined by German energy industry law as a "measuring device […], which reflects the real energy consumption of each user and the real time of consumption" [En10], represents a concrete modulation of the broader research ambition of the authors: to find out how big data and change/innovation processes in organizations can be handled, e.g. whether product managers of energy providers shall become more like data scientists or vice versa. We stress this topic to underline that in fact, most of the research done so far in science has been technologically initiated. Nevertheless, some authors (especially [KSB13] and [Go14]) have already pointed out the necessity for a more social-, business model- or user-oriented science approach within the field of EI to enable individuals and institutions to handle the potential opportunities (and through that, upcoming changes and innovations) presented by these technologies. In this context, we understand change management primarily in the classical view of Kotter, to prepare a corporation for new circumstances within the market [Ko12]. Furthermore, chances of relying on big data (and in this context, of smart metering) may also limit the potential for disruptive innovations with completely new solutions for problems [Ch13].

1.2 Research question

Within this overall frame, EI in general and smart meters in particular also offer an interesting research subject, combining the perspectives of technology, legal regulations and business models. Beyond our literature review, we are going to create a research agenda to point out which steps in science need to be taken to enrich the instruments of EI and IS due to this purpose.

The search term "big data" indicates a significant development in the Google Trend analysis (cf. figure 1, case sensitive search for all submitted queries worldwide between 2006 and 2015), which we introduce here as an additional indicator for the substantial growth of interest and publications on that topic. Google Trend normalizes all conducted searches between a scale from 0-100 by comparing a specific search query to the total amount of searches; therefore the Y-axis below can be understood as a general search trend over time.

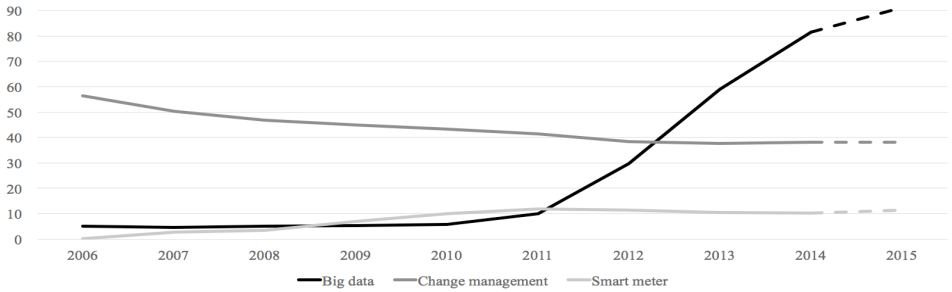


Figure 1: Google Trend indicates an increasing interest in big data

The chart also follows our above argument, that during the same timeframe the interest in the search query "change management" did not increase. Looking at the keyword "smart meter," a constant interest seems to be in place since 2011.

After presenting a wide scope of different perspectives on our topic, we can summarize our research with two specific questions:

RQ1: How can change and product innovation management enable the effectiveness (e.g. in terms of cost-savings or additional turnover) of big data initiatives?

RQ2: Which change and product innovation management enablers (e.g. communication, transformation strategy and so forth) can be applied for smart metering within EI?

To answer those two questions, we conduct a systematic literature review to investigate the status in EI/IS science and to draw a future research agenda. The paper is structured as follows: In section 2 the underlying research methodology is described. We present a quantitative analysis afterwards in section 3, including a discussion of open questions based on the studied papers. Section 4 provides a practical perspective on our topic by discussing selected cases from today's energy market in Europe. Finally, section 4 also points out which limitations we faced during our creation process and provides – as a main deliverable of this paper – a research agenda.

2 Research methodology

2.1 Systemic literature review and analyzed articles

Based on a systemic literature review process (cf. figure 2), the authors intend to follow a broadly accepted framework [Br09], which starts with a scope definition (see introduction), resulting in a conceptual keyword list. To focus only on results with high impact on the topic given, the following 16 search queries have been designed as Boolean match types: *big data implementation*, *big data and change management, big data and value chain, big data and business model, big data and decision making, big data and success factor, big data and customer relationship management, big data and*

product development, big data and service development, big data and organizational change, big data and hr, big data and CRM, big data and human resource, *big data and electricity, smart meter and value chain* and finally *smart meter and product innovation*.

In another step, the authors selected 23 A-ranked journals for research purposes, which was also based on a generally accepted journal ranking [WK08]. For the database search, EBSCO (Academic Search Premier, Business Source Premier and Regional Business News), SpringerLink, ScienceDirect and Google Scholar have been used. Additionally, we included the three top ranked IS conferences (ECIS, ICIS and WI) and entered the accompanying articles via AISeL with the same keyword-structure as used for the journal research.

Within the forward/backward search, our approach aimed on staying close to the research topic. Therefore the same quality premises for the initial search have been taken. Additionally, with the purpose of keeping a narrow perspective on the research subject, either the keyword "big data" or not predefined but impactful aspects must have been visible. In order to keep the quality of the source selection model described above, the criteria publication timeframe, journal respective conference working has also been used while executing the forward and backward literature search.

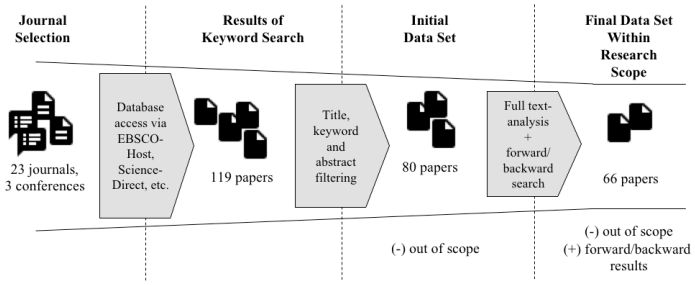


Figure 2: Funnel view of the research methodology

2.2 Related work

Within the research result we found three related articles which are summarized in tab. 1: One article provides a literature review about EI, pointing out that EI move beyond the solely technological aspects. The other two articles tried to answer the question, how big data initiatives can return value to an organization.

Title, Reference, Method	"Energy informatics. Current and future research focus." [Go14], review based approach
Publications reviewed	No explicit period; books, articles, proceedings
Research objective(s)	Creating an overview about the status quo in EI, especially intelligent energy saving systems and smart grids

Results	▪ Sensors and corresponding software are the key for smart grid technology and future developments ▪ EI shall include a market-oriented perspective, taking economic, psychological and sociological aspects into account – this can be achieved by focusing on systems and mechanisms to increase energy efficiency and by integrating renewable energy sources into the power grid
Title, Reference, Method	"Business intelligence and analytics: From big data to big impact" [CCS12], review based approach
Publications reviewed	Approx. 2002 - 2012, academic and industry publications
Research objective(s)	Reflecting the impact of data-related problems in corporations
Results	▪ Structured literature review of almost one decade research framework to identify emerging big data methods ▪ Analysis of applications, data, analytical methods and potential impact for five branches
Title, Reference, Method	"Creating value with business analytics in the supply chain" [SB13], case study
Publications reviewed	No explicit period; books, articles, proceedings
Research objective(s)	To use the resource-based view and dynamic capabilities for implementing business analytics into a company's value chain
Results	▪ Implementation framework for business analytics into an organization from a resource-based view ▪ Governance, change management, organizational learning, BA-enabled innovations and hybrid skills are important to gain value from big data

Tab. 1: Summary of related work

While the related papers focus either on a combination of the research perspectives big data and change/innovation or big data and smart metering, our paper will try to bring all three aspects together (cf. figure 3). As every related paper already offers a high value and well structured systematic for each viewpoint, our review is going to define a deep dive route through published articles and proceedings with a combined perspective on it.

3 Analysis of results

3.1 Research focuses of analyzed papers

The original search results of 119 articles have been reviewed based on their titles, keywords and abstracts in order to filter out those papers without any relation to our research. In the next step, the 80 remaining articles were analyzed in detail. Some

additional cases have been excluded within this step, as the content and applied methods in those articles did not correspond with the assumption based on keywords and abstracts; for all others, a significant fit to the given topic has been found and a forward-/backward search has been executed. All in all, this approach generated a data set of 66 articles, which will be examined in detail in the following sections. By filtering our result list of 66 articles by topic, a clear picture of the focus of each article can be generated. This indicates once again that big data in combination with change/innovation has already been the focus of previous research. Nevertheless, a combined view of smart metering and one or both of the other IS centric views show potential for further research activities.

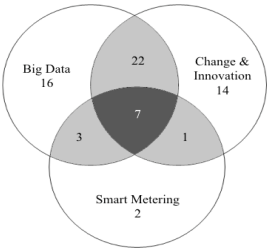


Figure 3: Overlapping and solitary research focuses found

3.2 Quantitative and chronological result analysis

The amount of published articles over time follows the trend of search queries on Google (cf. figure 4): At the time of producing this article, the 2015 level has already reached the total level for 2012. Almost 70% of the selected sources have published corresponding research between 2006 and today, distributed in a total of 15 journals respectively published at one of the three top ranked conferences.

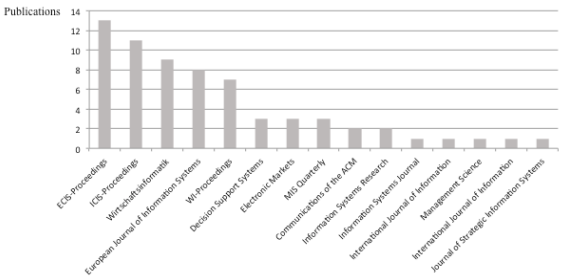


Figure 4: Accumulated result per journal/conferences (ranked by results per journal)

3.3 Research perspective of analyzed papers

To understand the meta-layer of each paper in a more detailed way, we selected four different perspectives based on an adapted PEST-analyze [FN86]: Papers with a clear

technical (we added mathematical) perspective formed one cluster, those with a clear economical perspective a second one. Articles with an organizational (we added social) view formed a third cluster. Finally, those articles with a broad perspective on the general orientation of their scientific discipline form the last cluster.

	2006	2008	2009	2010	2011	2012	2013	2014	2015
Technological/ Mathematical		1	1	1	1	2	2	4	4
Economical	1		2			2	6	3	2
Organizational/ Social	1	2	2	2			4	7	1
General Science						3	6	6	

Tab. 2: Research perspective of analyzed articles over time

Between 2006 and 2011, the majority of articles focused roughly on big data in terms of large and high frequency data-modulation within ERP projects, business analytics and intelligence and as well on change, innovation or implementation aspects. Looking on the general articles in our research scope, like [Go14b], [NM14] or [KSB13], it becomes clear that more and more, journal and conferences call for specific research according our research questions. While there has always been a technical or mathematical interest in our research topic, only latest research tries to provide a cross-disciplinary approach: For instance in [Ha15a] the authors analyze the digital transformation effects of primarily physical industries.

3.4 Applied research methods of analyzed papers

We analyzed the methods used by each paper to understand the development, status quo and possible gaps for a future research agenda. For this step, we followed a classification of research methods by Wilde and Hess [WH07]. The diagram below shows that empirical methods are used quite often in the analyzed articles, because papers like [FSN12] or [Wa13] selected a very specific problem to be solved with statistical models.

We would like to highlight the case studies in [KB14], where big data implementation into decision-making processes has been analyzed: The result, that the role definition of an enabler/integrator is necessary in the practical field correspondents with our finding of a research gap in science. In [Bu13a] the conducted interview goes more into detail, carving out that a mixed skill set of IT-tools, statistical education, but also a strong feeling for profitable business models will be required.

In terms of smart metering and the question how to utilize this from an economical perspective, two papers ([WBN15] and [AZ14]) offer initial studies into this direction. Under the economical aspect, general big data studies provide wider and more empirical research (e.g., in terms of supply chain benefits [SB13] or in the direction of data-driven business models [LG09], [OA13] and [Ha15a]).

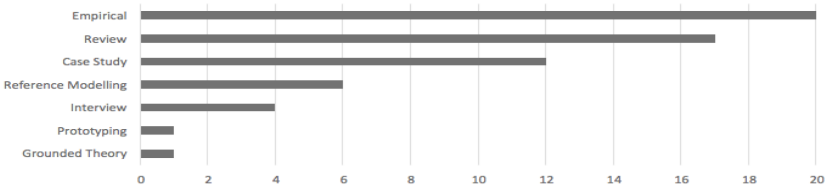


Figure 5: Research methods applied

The analysis is limited by the fact that not all authors specifically introduced their research methods. In those (minor) cases we tried to identify the chosen path and method of the authors by studying the presented approach and results. Nevertheless, a methodological test or validation was not in our scope.

3.5 Outcome: Problems to be solved

Following our used methodology from the previous two sections, we studied the open problems and research questions of each paper. We are going to highlight the main topics and representatives of each group in tab. 3.

Perspective	Problems to be solved	References
General science discipline	• Implementation of data-driven thinking into the different disciplines (EI/IS) • Further discussion of moral and ethical questions like data privacy necessary	[Bi13], [Bu13b], [Wo14]
	• Curriculum adaptations for a data-driven but also environmental-driven education • Combination of social, technological, data- and business-driven viewpoints	[Bi13], [AD14]
	• Finding a guiding viewpoint for EI specific articles • Implementing an interdisciplinary perspective into EI	[Re14], [KBK12], [Go14b], [KSB13]
Technological/ mathematical	• Data protection and privacy in terms of smart metering • General quality assurance in big data initiatives	[SK12], [ESB10], [LRI14], [HH14]
	• Potential of matching smart grids, smart meters, electric vehicles and other decentralized sources • Future data availability and effectiveness for steering of smart grids	[BF14], [SK12], [Kr15], [KW15], [KBK12]
Economical	• Value creating implementation of smart metering into energy retail organizations	[AZ14], [BWN12],

		[KBK12]
	• Future business models for data aggregators, collecting and using data from smart meters and offering their infrastructure for data processing to electricity retailers	[BF14]
	• Descriptive metrics not only regarding the change of one company, but also the change of an ecosystem of companies through digitalization	[Bh13]
	• Optimization of the intraday electricity purchase by predicting capacity and usage of renewable energies and electric vehicles	[BFN13]
	• Granting value added to consumers based on smart metering	[WBN15]
Organizational/ social	• Upcoming behavioral risks through big data	[Ol14]
	• Reduction of implementation costs of big data initiatives	[BH08], [By10]
	• Handling decision processes in organizations, which become more complex through big data	[Bu13a]
	• Cultural / language influences on skill profiles of big data experts	[DMV14]
	• In-depth case studies / expert interviews on digitalization of physical industries	[Ha15a]
	• Definition of role, skills and implementation plan of an analytical integrator as bridge between decision makers and data scientists	[KB14]
	• Motivating employees for creative use of data and stimulating data-driven thinking	[KLS14], [AD14]

Tab. 3: Overview of open problems and questions

4 Implications and outlook

4.1 Practical implications: Current trends in the field of smart metering

Based on a simplified approach of data triangulation [Fli11], we will add a practical view to the research we conducted. While the quantified review perspective before offered a structured path through related science of the last years, the chosen second perspective may also support timing and practical relevance of our topic.

As the lawmaker pushes smart metering, energy suppliers started already many initiatives to provide measuring devices for instance to private households. At this point we refer once again to Kotter's fundamental definition of change management, which

aims on preparing corporations for new market circumstances [Ko12]. Those market transformations are – based on the BITKOM association – already visible in the electricity branch and they will change the relationships between consumers and suppliers fundamentally, empowered by big data: The consumers' market position will be strengthened by transparency about their own consumption, while supplier and retailer will face a data-driven change in their business model [Bi13].

The consulting company PWC identified key steps to handle the transition from the classical energy- to a modern service-provider, which summarizes our findings with regards to RQ1: As a primarily step, the overall business model and strategy needs to be adapted due to the digital transformation of the branch. Afterwards, the organizational change has to be conducted. This will take effect in recruiting and job scopes for example, because new skills are necessary to work in a service-oriented company instead of working for a classical energy supplier. In a third step the implementation into the core processes of the company must take place, i.e. by creating a service center [PW13].

Vestas Wind can be taken as a possible path for energy enterprises to react on RQ1: The energy supply company was able to identify the initial problem (wind forecasting) as a big data challenge. By gathering weather data from internal and external sources, a weather library with 24 petabytes of data has been created. Referring to tab. 3 (especially [BFN13]), Vestas faced as well the challenge how to monetize the data. Two approaches were implemented: Earning savings within the intraday market through better wind forecasts and a change in the investment strategy for new wind turbines by using more precise geographical prediction of the wind flow. This example underlines our economical and organizational outcomes of tab. 3: Although the technical enhancement seems to be a fundamental step, Vestas' management faced the barrier of a value-oriented utilization of the additional data [IB11].

This barrier, which can be described more precisely as the gap between data science and corporate decision-making, is also an outcome of Kowalczyk's and Buxmann's research and puts an additional aspect to RQ1 [KB14]: In a multiple case study, big data & BI experts have been interviewed about past and running big data initiatives. In general, an effective integration of data-centric decisions requires domain specific knowledge of the decision maker on the one hand and the communicational expertise, analytical understanding and visualization skills of an integrator on the other hand – as also discussed in section 3.4. The definition of that role from an integrated ES/IS perspective might also form a starting point for further research.

Beside that interpersonal or cultural aspect, another case study about the use of sensor data discovered a process related cornerstone for big data implementation: Organizations must prepare their own processes for insights which might come up through big data. In the given case study, people were able to mark potholes on the rode in a mobile device app. The GPS-based data grew so fast, that the pilot city Boston was not able to handle the following processes [OL13]. Applying this lesson learnt to RQ2 and to the energy market, where retailers still struggle to predict consumers' energy consumption with classical volatility models – i.e. without sensor based and individual data [FSN12] –

retailers' business models and core processes must get prepared for a smart meter and data-driven direction.

4.2 Research implications: Elements of a research agenda and conclusion

Returning to our two research questions, we asked how big data could generally, and in terms of smart metering more specifically, enable change and innovation. By presenting a selection of current real world cases it became clear, that those questions are posed at the right time and – following the interdisciplinary approach – into the right direction. Furthermore we outlined a possible implementation and change management path for big data driven products in the electricity branch.

As every research paper has its limitations, we would like to point out that by selecting a cross-disciplinary perspective, not every reviewed paper can offer a combination of all three perspectives (big data, change/innovation and smart metering). Therefore, we suggest primarily further research to identify the scope of the problem. Another limitation arises through the practical real world cases which we included into section 4.1, which where not purely based on academic literature but also on industrial publications.

The literature review that we conducted within this paper shall be followed by further in-depth case studies. In that step, as well as during the empirical study, other branches with experience in data- and sensor-driven businesses might also be reflected. For instance, automakers are facing a quite similar discussion regarding connected cars right now. During the second step, the conceptual phase, a reference model and a business maturity model shall be developed which can be tested and improved empirically in the third step [BKP09]. In regular frequencies the previous research shall be reflected – e.g., action research can deliver additional insights for this or new case studies can be conducted.

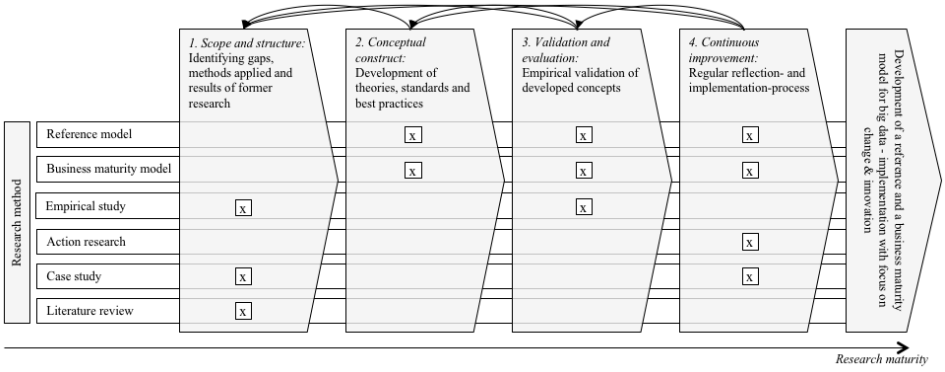


Figure 6: Elements of a research agenda, based on [MF09]

References

The complete reference list, including the 66 articles analyzed in our literature review, is available on: http://goo.gl/AZdSIs

[Bi13] BITKOM: Management von Big-Data-Projekten. Leitfaden. Available on https://www.bitkom.org/files/documents/LF_big_data2013_web.pdf, downloaded on the 31st of March 2015.

[BKP09] Becker, J.; Knackstedt, R.; Pöppelbuß, J.: Developing Maturity Models for IT Management. Business & Information Systems Engineering, 1(3), P. 213–222, 2009.

[Br09] Von Brocke, J. et.al.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. ECIS-Proceedings, Verona 2009.

[Ch13] Christensen, C.M.: The innovator's dilemma. When new technologies cause great firms to fail. Boston, 2013.

[En10] Energiewirtschaftsgesetz – EnWG §21c: Gesetz über die Elektrizitäts- und Gasversorgung. Einbau von Messsystemen, 2010.

[Eu06] European Union: Directive 2006/32/EC of the European Parliament and of the council of 5 April 2006 on energy end-use efficiency and energy services and repealing Council Directive 93/76/EEC, 2006.

[Fli11] Flick, U.: Triangulation. Eine Einführung. Wiesbaden, 3rd Edition, 2011.

[IB11] IBM: Vestas. Turning climate into capital with big data. Available on http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=PM&subtype=AB&htmlfid=IMC14702USEN#loaded, downloaded on the 6th of April 2015.

[Ko12] Kotter, J.P.: Leading Change. Massachusetts, 2012.

[MF09] Martens, B.; Teuteberg, F.: Why Risk Management Matters in IT Outsourcing - A Systematic Literature Review and Elements of a Research Agenda. ECIS-Proceedings, Verona 2009.

[OL13] O'Leary: Exploiting big data from mobile device sensor-based apps: Challenges and benefits. MIS Quarterly Executive, 12(4), P. 179–187, 2013.

[PW13] Smart Metering. Intelligente Messsysteme für die Energienetze von morgen. Available on http://blogs.pwc.de/auf-ein-watt/files/2013/08/Smart_Metering-Intelligente_Messsysteme_f%C3%BCr_die_Energienetze_von_morgen.pdf, downloaded on the 31st of March 2015.

[TS13] T-Systems International GmbH: Energie durch Big Data. Sind Europas Energieversorger für das Datenzeitalter gerüstet? 2013.

[VB09] Von Brocke, J.; Simons, A.; Niehaves, B. et.al.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. Retrieved February 25, 2015.

[WH07] Wilde, T.; Hess, T. (2007). Forschungsmethoden der Wirtschaftsinformatik – Eine empirische Untersuchung. Wirtschaftsinformatik, 49 (4), P. 280-287, 2007.

[WK08] WKWI: WI-Orientierungslisten. Wirtschaftsinformatik 50 (2), P. 155-163, 2008.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Modelling the Performance, Energy Consumption and Efficiency of Data Management Systems

Raik Niemann[1] and Todor Ivanov[2]

Abstract: Many approaches, proposals and surveys to evaluate, measure and optimize the energy efficiency of data management systems were made during the last couple of years stressing the importance of this research field.

From the practitioner's perspective, for example developers and end users of data management systems, there is a big number of factors that influence the performance and energy efficiency of a particular data management system. For example, the replacement of hardware components or the surrounding operating system can have a significant impact. Both developers and end users put much effort into finding performance "bottlenecks", better hardware resource utilization and configurations. Besides, when it comes to a scale-out scenario, end users often face the situation to find a hardware configuration that offers both a reasonable performance and energy consumption, i.e. a resource planning.

Multiple evidence presented in this paper suggest that a model to simulate both the performance and energy consumption of a data management system is feasible. This allows to evaluate the energy efficiency by running a simulation rather than by performing real experiments. Compared to traditional ways, for example regression and compatibility tests, simulation runs are intended to drastically reduce the investments, both in time and hardware.

In this paper, both the reasons to create such simulation model as well as the characteristics and benefits of its fundamental design were discussed.

Keywords: Performance, Energy efficiency, Data management system

1 Introduction

The incentive to deal with the energy efficiency arises in many fields. When the number of incorporated components to handle a problem reaches a specific limit, the components themselves become a cost factor. Examples for such fields are data management and processing systems as well as sensor and mobile device networks.

To illustrate this issue, one can consider the development of data management systems in the last couple of years. The amount of information that is being created worldwide increases exponentially. The traditional data management and processing systems are not able to store and to process this amount efficiently. This leads to all the efforts that can

[1] Institute of Information Systems, Departement of Multimedial Information Systems, University of Applied Science Hof, Hof (Germany), raik.niemann@iisys.de
[2] Chair for Database and Information Systems, Institute for Informatics and Mathematics, Goethe University Frankfurt, Frankfurt am Main (Germany), todor@dbis.cs.uni-frankfurt.de

be summarized under the term “Big Data”. Examples are the appearance of *NoSQL* data management systems or data processing frameworks like *Hadoop*.

Although those systems allow the storage and processing of huge data amounts more efficiently, operators are still challenged. For example, there is a high number of combinations regarding the selection of operating systems, data management systems and hardware configurations. When it comes to a scale-out scenario or resource planning according to changed demands and requirements, tradeoffs are existing. The term “green computing” combines all efforts to find those tradeoffs.

It has been several years since the *Claramont report* [AAB+08] showed the directions, challenges and potential of “green computing”. The intention was to stimulate the research of new architectures, use cases and implementations of data management systems. Although some of them were realized, the reality shows that energy efficiency optimizations of data management systems are ambigous as outlined by *Wang et al.* [WFXS11].

The remaining paper is organized as follows: Section 2 briefly describes the calculation and methodology of the energy efficiency. Section 3 discusses issues with respect to the evaluation of the performance and energy consumption of data management systems. Section 4 addresses these issues and shows –as a solution– that a model for the simulation of both the performance and energy consumption is practical. Section 5 describes the fundamental design, the model and its key benefits whereas Section 6 presents the results of our first simulation models. Section 7 summarizes the paper.

2 Evaluation of the Energy Efficiency

There is a well known equation to calculate the energy efficiency *EE* of a data management system ([WFXS11], [XTW10]):

$$EE = \frac{W_{DB}}{W_e} = \frac{Perf \cdot t}{P \cdot t} = \frac{Perf}{P} \tag{1}$$

where W_{DB} is the useful work the data management system did and W_e the electrical work it consumed. P denotes the mean effective power. As W_{DB} is something arbitrary, it can not be expressed with a physical unit. In general W_{DB} is the equivalent of the mean performance $Perf$ over the time t.

As Equation 1 states, it is essential for the calculation of the energy efficiency to measure both the mean performance and the energy consumption. The former is given by a test methodology and the latter by a definition of the test apparatus. For comparability reasons, standards have been developed for both.

For instance, to evaluate the performance of a data management system, a benchmark implementing one or more real use cases, is run on top of it. There are widely accepted benchmarks for all kinds of data management systems. All these benchmarks are similar in that they report as a performance metric either the response time or the throughput. In

general the response time is determined by executing a specific workload and measuring the time period from the initial access of the data management system until the end of the execution. This can be done simultaneously with a higher number of clients to evaluate the throughput.

For measuring the energy consumption while performing a benchmark, one or more energy consumption measurement devices are used. The test apparatus provides a description on how to measure the energy consumption. It also provides a list with the system components that should be measured.

3 Issues in the Evaluation of Energy Efficiency

There are several issues when it comes to the evaluation and optimization of the energy efficiency with respect to data management systems. Equation 1 states that any optimization follows either the increase of the performance or the reduction of the energy consumption.

For end users and developers of data management systems the optimization direction is important and still not clear [WFXS11]. Since the common way to evaluate and calculate the energy efficiency is by performing benchmarks while measuring the energy consumption, both groups have to invest into time (for the benchmark) and hardware components (for the benchmark running on different combinations of hardware components). As a result, all efforts lead to the *comparability* of data management systems in terms of performance, energy consumption and finally energy efficiency. The ability to compare two different data management systems or two different versions of the same one allows the decision making and the resource planning. For instance, developers of data management systems need comparability as the basis for regression tests. Similarly, end users need comparability to choose the more performant and energy efficient data management system for a given use case.

A general comparability of all known data management systems is impossible because this requires a general benchmark that covers all imaginable use cases and a common access interface to the data management systems in order to evaluate their performance. Additionally such a benchmark also requires a general methodology and test apparatus to measure the energy consumption for all possible combinations of hardware components.

Despite of that, it is feasible to compare data management systems that have a common access interface or a common domain. There are benchmarks for most domains that standardize the workloads to evaluate the performance of the tested data management systems. However, these benchmarks are not consistent regarding the reported metrics which makes the result comparision more difficult. A recent survey on benchmarks for *Big Data* [QZ13] extensively reviews the current metrics and discusses some of the characteristics and challenges they should address. An important point that the authors make is the need of a complete, end-to-end benchmarking suite, including both component-based and application-oriented benchmarks along with critical metrics such as energy consumption. On the contrary, more general metrics which are independent of workload type or based on processor micro-architecture characteristics are reported. Examples are Cycles per In-

structions (CPI), first level data cache misses per 1000 instructions (L1 MPKI) and last level cache (LLC) miss ratio, presented in [XYB+13] and [DKL+13]. Finally, important new metrics like data processed per second and data per Joule are proposed by [LZJ+12] to better measure data processing and energy consumption.

In contrast to the evaluation of the performance, the methodology to measure the consumed electrical energy according to Equation 1 is not consistent. In particular we refer to the following issues:

- There are only two major standards that are widely accepted: *SPECpower* [(SP15] and *TPC-Energy* [(TP15]. In contrast to *TPC-Energy* which standardizes an additional metric for their *TPC-(C,E,H,DS)* benchmark, *SPECpower* is a more general way to relate the performance to the energy consumption.
- The set of components to be measured is not consistent. *TPC-Energy* as well as *SPECpower* define the components of a system under test (SUT) but some of them are only optional. However, the reality shows that only the measurement of the overall power consumption matters. This is based on the fact that both *TPC-Energy* and *SPECpower* require the usage of a separate software that is responsible for the entire benchmark execution and also to measure the energy consumption of the incorporated devices. In fact, this software supports only a couple of measurement devices.
- Another important aspect is the inconsistent definition of the time frame that should be measured. *TPC-Energy* defines this as the time period that it takes to execute the benchmark. *SPECpower* extends this time period with short time periods before and after the benchmark (so called “ramp-up”). The reason for this is to reflect the preparation of the benchmark.

4 Reasons for a Model

In our opinion all approaches that were developed in the last decade towards the optimization of the energy efficiency in data management systems showed the same empirical progress and semantics as the ones known from the behavioral science.

This means that three stages are required to fully understand a topic. In the first stage all factors that could have a potential influence are gathered. In the second stage the influence factors are tested individually to estimate their impact. In the final third stage a model abstracts all findings and allows the deduction of new knowledge.

With respect to the topic of this paper it is our belief that enough research results were published to finally form a model. To prove our point we surveyed prominent publications and approaches.

The first research concerning the factors that could have a potential influence on the energy efficiency were evaluated and presented by *Xu et al.* [XTW10], *Wang et al.* [WFXS11] and *Tsirogiannis et al.* [THS10]. Although they focused on relational data management systems, important influence factors were identified and first measurement methodologies

were proposed. In addition to this the assumption was verified that the optimization of single influence factors have a significant impact on the energy efficiency. *NoSQL* data management systems became very popular, especially in conjunction with the *Hadoop* ecosystem. This introduced additional potential influence factors as presented by *Lang et al.* [LKP11], *Abadi* [Aba12], *Rabl et al.* [RGVS+12] and *Baru et al.* [BBN+13].

The impact of the identified factors was evaluated by executing benchmarks for fixed combinations of those factors, as reported for example by *Vasic et al.* [VBSK09], *Lang and Patel* [LP09] and *Poess et al.* [PNV+10]. Prominent approaches modified existing data management systems or their test apparatus to optimize either the performance (response time and throughput) or the energy consumption of the system. *Hindman et al.* [HKZ+11] tried to get existing data management systems aware of the energy consumption which is in return just another aspect of optimizing for the performance. This was done by combining existing heuristics like indices and query planner statistics with energy consumption metrics. Another approach was to take energy efficiency into consideration when the data management software is in its planning and implementation stage (so called *software performance engineering* as described in [WFP07] and [DM02]). A prominent example for software performance engineering is *F1* [SOE+12], a relational data management system developed by *Google* addressing the energy efficiency and performance as key requirements.

In addition to the given publications from the academia, even more white papers were released to the public by the developers of modern data management systems. Those white papers present experimental performance results as well as architectural insights (for example [Dat14a] and [Dat14b] for *Cassandra*).

5 Model Characteristics

As implied by Equation 1, a model has to simultaneously consider both the performance and energy consumption aspects of a system. To achieve this, both aspects need to be abstracted and generalized. In particular this means to find similarities that data management systems, benchmarks and benchmark methodologies have in common.

First, we studied the similarities of prominent data management systems. The result was that they only share the basic fundamental architecture ([WFXS11], [WV01]). In principle they offer an interface to save and relate facts or information of a specific domain with respect to the *ACID* properties (atomicity, consistency, isolation and durability). They also offer an interface to query the data in an efficient and performant manner by using heuristics, for example by using different indexing mechanisms.

Second, we studied major benchmarks in order to find their similarities. To get an overview how the major benchmarks define their performance metric, we surveyed them according to their reported metric. A summary of metrics is presented in Table 1. It is noticeable that the dominant metric is either throughput or response time. The latter can be expressed as throughput if the queried data amount is known. Next, we reconsidered all approaches given in the previous sections as well as the similarities mentioned above to find a model

that fully supports all aspects. We came to the conclusion that a basic data flow model is the most promising solution.

Benchmark	Category and metric
TPC-E	OLTP processing workload of a brokerage firm Transactions per second
TPC-C	OLTP processing workload for heavy transactions Transactions per minute
TPC-H	OLAP decision support workload for large sets of data Query response time
TPC-DS	Like TPC-H but with more I/O and CPU load Query response time
TPC-VMS	Additional methodology for virtualized TPC-(E,C,H,DS) Summarized throughput
BigData benchmark	OLAP benchmark for MPP architectures [AMP14] Query response time
BigDataBench	Compound OLTP and OLAP benchmark for various architectures [WZL+14] Throughput
HiBench	OLAP benchmark for *Hadoop*/*Hive* [HHD+10] Throughput and response time
BigBench	TPC-DS inspired benchmark [GRH+13] Throughput (queries per hour)
LinkBench	Social graph database benchmark [APBC13] Query response times
YCSB	Framework to benchmark cloud services, e.g. DBMS' [CST+10] Operations per second
XMark	Benchmark for XML operations in a data management system [SWK+02] Query response time
Cloudsuite	Compound benchmark for datacenter operations [FAK+12] (undecided)
TPCx-HS	Benchmark to stress *Hadoop* configurations Response time and throughput

Tab. 1: Surveyed data management system benchmarks with their metrics

The main reason for a general model is the fact that the data flow is fundamental and valid for all types of software that process data in a structured manner such as data management systems which share the same architecture. Furthermore, a model that is specifically designed to reflect the common data flow within a data management system has also the advantage that performance can be simulated and predicted. *Bontempi and Kruijtzer* [BK02], *Kounev et al.* [KGS08] and *Nambiar and Poess* [NP10] showed that it is possible to simulate and precisely predict the performance of a system.

Additionally the amount or structure of the input data for the data flow model does not matter. This is most useful with regard to benchmarks. Since all the surveyed benchmarks generate a predefined amount of domain specific data, this data can be used as input for the model. The advantage of this approach is that the simulation results can be directly

compared to experimental results. The surveyed benchmarks have also in common that they focus on the evaluation of either the response time or the throughput. With respect to the data flow model both metrics can be measured and reported.

Another reason for forming a model based on the data flow is the possibility to use additional metrics. According to *Dietrich et al.* [DBCRW92], a data flow model that includes all components required for processing data could also consider all hardware components. When it comes to the simulation of the data flow, an additional metric like the mean power consumption per simulated hardware component allows to calculate the overall power consumption.

In a nutshell, a data flow model whose components reflect the common architecture of data management systems as well as all the incorporated processing hardware is able to simulate the data flow that was previously created by a benchmark. In conjunction with an additional metric like the mean power consumption per component, the simulation of the overall power consumption is possible. Both the performance simulation results and the overall power consumption allow the calculation of the energy efficiency defined in Equation 1.

The key benefit of this model is the significant reduction in time and hardware investments. As outlined in Section 2 the common way to evaluate both the performance and energy consumption aspects is to execute a benchmark for comparability purposes. This needs to be done on different hardware configurations for a holistic evaluation of the energy efficiency. Simulating those aspects on the basis of the model would be faster without the need to invest into the desired hardware configuration.

From the practitioner's perspective, both developers and end users of data management systems would benefit from the proposed model. A variety of use cases are possible. They have all in common that only three basic simulation parameters are required: the data management system, the hardware configuration and the amount or domain of the data to be processed. Fixing two of this three parameters while varying the remaining one allows to simulate the effects on the performance and energy consumption.

For example, it enables developers to simulate the effects of possible implementation changes in the architecture of a data management system such as the replacement of the data storage layout or indices. This replacement can be made in the model instead of the software implementation. A simulation with a given amount of data and hardware configuration demonstrates the effects on the performance without having the architectural change actually implemented and evaluated by performing a benchmark. Moreover, the usual process of doing regression tests can be omitted.

In contrast to developers, end users of existing data management systems like datacenter operators are more interested in resource planning or scale-out scenarios. Given an existing data management system and hardware configuration, the effects on the performance and energy consumption can be simulated while changing the amount of data to be processed. The results can be used for further simulations, for example to find a hardware configuration that is more performant and energy-saving at the same time. Similarly, the most

energy efficient data management system can be identified by simulation if the hardware configuration and the data size are known in advance.

6 Model Implementation and First Results

In [NI15], the data flow of a data management system was implemented by using *Queued Petri Nets* (QPN[1]). This model is intended to reflect the data flow of a data management system among all components that could have an impact on the performance. This includes the surrounding operating system, technical components (main memory, mass storage and CPU) and the common components of a data management system (access control, query plan generation and optimization, query plan isolation and query plan execution).

The basic model was evaluated and proven by simulating two prominent database benchmarks (*TPC-H* and *StarSchema Benchmark*). The simulation runs reported the simulated query response time and statistical values for all components of the model, for example the throughput of transitioned marks per simulated second. The throughput is important when it comes to the calculation of the simulated energy consumption. The simulated query execution times, energy consumption and efficiency values were compared with the experimental ones that were presented in [NKZG13]. On average the experimental and simulated values for the *TPC-H* benchmark differ by 36.55 percent for the query execution times and by 21.91 percent for the energy efficiency, respectively. For the *StarSchema Benchmark*, the experimental and simulated values differ by 19.77 percent for the query execution times and by 20.77 percent for the energy efficiency, respectively.

However, the usage of the basic model is limited because simulation runs aim to reflect the data flow of a data management system running on a single server machine. As a result, it only allows to simulate and to study the effects of vertical scale-out scenarios.

To simulate the data flow of a data management system that is distributed over several machines, it was necessary to introduce a new QPN model that enables the existing models to exchange data in the form of marks. In other words, network functionality was added to the basic models in [Nie15]. Simulation runs using the enhanced models in combination with the new introduced model allow additionally to observe the effects of horizontal scale-out scenarios. The enhanced models are able to simulate any computer networks. This means that the topology, the data transmission bandwidth and performance can be taken into consideration.

To prove the enhanced models, experiments on a bladecenter with seven uniform blades were performed. The experiments used the *Yahoo Cloud Serving Benchmark* (YCSB, [CST+10]) on the the distributed data management system *Cassandra*. The enhanced models were set up and adjusted to reflect both the used real test apparatus and the benchmark. A comparison between the real and simulated experimental results showed a difference of 20.07 percent on average for the used YCSB workloads and 47.25 percent on average for

[1] A QPN combines Colored Petri Nets (CPNs), Generalized Stochastic Petri Nets (GSPNs) and queuing principles as well as scheduling strategies to the places. The formal definition for a QPN is fairly extensive and can be found in [KB03].

the energy efficiency, respectively. Note that to the best of our knowledge there is no other QPN model that is able to simulate the performance and energy consumption for the tested benchmarks. Therefore a comparison of the accuracy is infeasible.

In contrast to the experiments using the basic model mentioned before, the YCSB benchmark in combination with *Cassandra* turned out to be less complex including fewer transition rule and mark sets. This enabled the simulation of the workloads to finish in 3 minutes and 46 seconds on average to get a simulated response time of one YCSB workload. Compared to the experimental workload response times, this is nearly 8000 times faster.

The simulation model allowed to perform additional runs for *Cassandra* clusters having more than seven nodes. As a result, the simulation results for a given use case (fixed data size of 300 GByte to be stored in the *Cassandra* cluster) revealed that the optimal number of nodes is 10. This means that the performance and energy consumption gains fell under 10 percent with every subsequently added *Cassandra* node.

Further simulation runs showed a turning point for a *Cassandra* cluster having 25 nodes. At this point the simulated communication between the cluster nodes had an impact on the performance. As a result, the performance decreased by nearly 2.1 percent with every subsequently added *Cassandra* node. Note that these additional simulation results can not be validated with real experiments since the used test apparatus had only seven nodes.

7 Conclusion

In our opinion there is enough academic research and experimental results to form a model that is capable of simulating and predicting both the performance and the energy consumption of a data management system. This allows to calculate the energy efficiency in order to optimize it.

Based on our survey, multiple reasons motivating why such a model is feasible were outlined. Next we introduced the fundamental components and characteristics of the model, together with its key benefits and advantages. In [NI15] and [Nie15], we presented an approach to simulate the data flow of three prominent benchmarks (*TPC-H*, *StarSchema* and *Yahoo Cloud Serving Benchmark*) for two popular data management systems (*PostgreSQL* and *Cassandra*). The data flow was modelled using *Queued Petri Nets*. The executed simulation runs allowed to predict the response time of the benchmark workloads and to calculate the overall energy consumption for the simulated hardware components.

The simulation results were compared with experimental ones and showed a difference of 25.46 percent on average for the query or workload response times and 29.97 percent on average for the energy efficiency, respectively. The simulation runs were up to 8.000 times faster when compared to the real benchmark executions on the test servers. In [Nie15], additional simulation runs with changed hardware configuration revealed the turning point between the performance and energy consumption for a *Cassandra* cluster. In this way the most energy efficient hardware configuration for a given use case was found.

References

[AAB+08] Rakesh Agrawal, Anastasia Ailamaki, Philip A Bernstein, Eric A Brewer, Michael J Carey, Surajit Chaudhuri, AnHai Doan, Daniela Florescu, Michael J Franklin, Hector Garcia-Molina, et al. The Claremont report on database research. *ACM SIGMOD Record*, 37(3):9–19, 2008.

[Aba12] Daniel Abadi. Consistency tradeoffs in modern distributed database system design: CAP is only part of the story. *Computer*, 45(2):37–42, 2012.

[AMP14] AMPLab. BigData benchmark, 2014. https://amplab.cs.berkeley.edu/benchmark/.

[APBC13] Timothy G Armstrong, Vamsi Ponnekanti, Dhruba Borthakur, and Mark Callaghan. Linkbench: a database benchmark based on the facebook social graph. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, pages 1185–1196. ACM, 2013.

[BBN+13] Chaitanya Baru, Milind Bhandarkar, Raghunath Nambiar, Meikel Poess, and Tilmann Rabl. Benchmarking Big Data Systems and the BigData Top100 List. *Big Data*, 1(1):60–64, 2013.

[BK02] Gianluca Bontempi and Wido Kruijtzer. A data analysis method for software performance prediction. In *Proceedings of the conference on Design, automation and test in Europe*, page 971. IEEE Computer Society, 2002.

[CST+10] Brian F Cooper, Adam Silberstein, Erwin Tam, Raghu Ramakrishnan, and Russell Sears. Benchmarking cloud serving systems with YCSB. In *Proceedings of the 1st ACM symposium on Cloud computing*, pages 143–154. ACM, 2010.

[Dat14a] DataStax. Apache Cassandra 2.0 Documentation, 2014. http://docs.datastax.com/en/cassandra/2.0/pdf/cassandra20.pdf.

[Dat14b] DataStax. DataStax Enterprise Reference Architecture, 2014. http://www.datastax.com/wpcontent/uploads/2014/01/WPDataStaxEnterpriseReference-Architecture.pdf.

[DBCRW92] Suzanne W. Dietrich, M Brown, Enrique Cortes-Rello, and S Wunderlin. A practitioner's introduction to database performance benchmarks and measurements. *The Computer Journal*, 35(4):322–331, 1992.

[DKL+13] Martin Dimitrov, Karthik Kumar, Patrick Lu, Vish Viswanathan, and Thomas Willhalm. Memory system characterization of big data workloads. In *Big Data, 2013 IEEE International Conference on*, pages 15–22. IEEE, 2013.

[DM02] Elizabeth D Dolan and Jorge J More. Benchmarking optimization software with performance profiles. *Mathematical programming*, 91(2):201–213, 2002.

[FAK+12] Michael Ferdman, Almutaz Adileh, Onur Kocberber, Stavros Volos, Mohammad Alisafaee, Djordje Jevdjic, Cansu Kaynak, Adrian Daniel Popescu, Anastasia Ailamaki, and Babak Falsafi. Clearing the clouds: a study of emerging scale-out workloads on modern hardware. In *Proceedings of the seventeenth international conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS '12, pages 37–48, New York, NY, USA, 2012. ACM.

[GRH+13] Ahmad Ghazal, Tilmann Rabl, Minqing Hu, Francois Raab, Meikel Poess, Alain Crolotte, and Hans-Arno Jacobsen. BigBench: towards an industry standard benchmark for big data analytics. In *Proceedings of the 2013 ACM SIGMOD international conference on Management of data*, pages 1197–1208. ACM, 2013.

[HHD+10] Shengsheng Huang, Jie Huang, Jinquan Dai, Tao Xie, and Bo Huang. The HiBench benchmark suite: Characterization of the MapReduce-based data analysis. In *Data Engineering Workshops (ICDEW), 2010 IEEE 26th International Conference on*, pages 41–51. IEEE, 2010.

[HKZ+11] Benjamin Hindman, Andy Konwinski, Matei Zaharia, Ali Ghodsi, Anthony D Joseph, Randy H Katz, Scott Shenker, and Ion Stoica. Mesos: A Platform for Fine-Grained Resource Sharing in the Data Center. In *NSDI*, volume 11, pages 22–22, 2011.

[KB03] Samuel Kounev and Alejandro Buchmann. Performance modelling of distributed e-business applications using Queuing Petri Nets. In *Performance Analysis of Systems and Software, 2003. ISPASS. 2003 IEEE International Symposium on*, pages 143–155. IEEE, 2003.

[KGS08] Samuel Kounev, Ian Gorton, and Kai Sachs, editors. *Performance Evaluation: Metrics, Models and Benchmarks*, volume 5119, Darmstadt, Germany, June 2008. SPEC International Performance Evaluation Workshop, Springer.

[LKP11] Willis Lang, Ramakrishnan Kandhan, and Jignesh M Patel. Rethinking Query Processing for Energy Efficiency: Slowing Down to Win the Race. *IEEE Data Eng. Bull.*, 34(1):12–23, 2011.

[LP09] Willis Lang and Jignesh Patel. Towards eco-friendly database management systems. *arXiv preprint arXiv:0909.1767*, 2009.

[LZJ+12] Chunjie Luo, Jianfeng Zhan, Zhen Jia, Lei Wang, Gang Lu, Lixin Zhang, Cheng-Zhong Xu, and Ninghui Sun. CloudRank-D: benchmarking and ranking cloud computing systems for data processing applications. *Frontiers of Computer Science*, 6(4):347–362, 2012.

[NI15] Raik Niemann and Todor Ivanov. Evaluating the Energy Efficiency of Data Management Systems. In *4th International Workshop on Green and Sustainable Software, ICSE 2015*. IEEE, 2015.

[Nie15] Raik Niemann. Evaluating the Performance and Energy Consumption of Distributed Data Management Systems. In *1st Green in Software Engineering Workshop, ICGSE 2015*. IEEE, 2015.

[NKZG13] Raik Niemann, Nikolaos Korfiatis, Roberto Zicari, and Richard Göbel. Evaluating the Energy Efficiency of OLTP Operations. In *Availability, Reliability, and Security in Information Systems and HCI*, pages 28–43. Springer, 2013.

[NP10] Raghunath Nambiar and Meikel Poess, editors. *Performance Evaluation, Measurement and Characterization of Complex Systems*, volume 6417, Singapore, September 2010. Second TPC Technology Conference, TPCTC, Springer.

[PNV+10] Meikel Poess, Raghunath Othayoth Nambiar, Kushagra Vaid, John M Stephens Jr, Karl Huppler, and Evan Haines. Energy benchmarks: a detailed analysis. In *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking*, pages 131–140. ACM, 2010.

[QZ13] Xiongpai Qin and Xiaoyun Zhou. A Survey on Benchmarks for Big Data and Some More Considerations. In *Intelligent Data Engineering and Automated Learning–IDEAL 2013*, pages 619–627. Springer, 2013.

[RGVS+12] Tilmann Rabl, Sergio Gómez-Villamor, Mohammad Sadoghi, Victor Muntés-Mulero, Hans-Arno Jacobsen, and Serge Mankovskii. Solving big data challenges for enterprise application performance management. *Proceedings of the VLDB Endowment*, 5(12):1724–1735, 2012.

[SOE+12] Jeff Shute, Mircea Oancea, Stephan Ellner, Ben Handy, Eric Rollins, Bart Samwel, Radek Vingralek, Chad Whipkey, Xin Chen, Beat Jegerlehner, Kyle Littlefield, and Phoenix Tong. F1 - The Fault-Tolerant Distributed RDBMS Supporting Google's Ad Business. In *SIGMOD*, 2012. Talk given at SIGMOD 2012.

[(SP15] Standard Performance Evaluation Corporation (SPEC). SPEC Power and Performance Methodology, 2015. https://www.spec.org/power/docs/SPEC-Power_and-_Performance _Methodology.pdf.

[SWK+02] Albrecht Schmidt, Florian Waas, Martin Kersten, Michael J Carey, Ioana Manolescu, and Ralph Busse. XMark: A benchmark for XML data management. In *Proceedings of the 28th international conference on Very Large Data Bases*, pages 974–985. VLDB Endowment, 2002.

[THS10] Dimitris Tsirogiannis, Stavros Harizopoulos, and Mehul A Shah. Analyzing the energy efficiency of a database server. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pages 231–242. ACM, 2010.

[(TP15] Transaction Processing Performance Council (TPC). TPC Energy Specification, 2015. http://www.tpc.org/TPC_Documents_Current_Versions/pdf/TPC Energy Specification 1.5.0.pdf.

[VBSK09] Nedeljko Vasić, Martin Barisits, Vincent Salzgeber, and Dejan Kostic. Making cluster applications energy-aware. In *Proceedings of the 1st workshop on Automated control for datacenters and clouds*, pages 37–42. ACM, 2009.

[WFP07] Murray Woodside, Greg Franks, and Dorina C Petriu. The future of software performance engineering. In *Future of Software Engineering, 2007. FOSE'07*, pages 171–187. IEEE, 2007.

[WFXS11] Jun Wang, Ling Feng, Wenwei Xue, and Zhanjiang Song. A survey on energy-efficient data management. *ACM SIGMOD Record*, 40(2):17–23, 2011.

[WV01] Gerhard Weikum and Gottfried Vossen. *Transactional information systems: theory, algorithms, and the practice of concurrency control and recovery*. Elsevier, 2001.

[WZL+14] Lei Wang, Jianfeng Zhan, Chunjie Luo, Yuqing Zhu, Qiang Yang, Yongqiang He, Wanling Gao, Zhen Jia, Yingjie Shi, Shujie Zhang, et al. Bigdatabench: a big data benchmark suite from internet services. In *High Performance Computer Architecture (HPCA), 2014 IEEE 20th International Symposium on*, pages 488–499. IEEE, 2014.

[XTW10] Zichen Xu, Yi-Cheng Tu, and Xiaorui Wang. Exploring power-performance tradeoffs in database systems. In *Data Engineering (ICDE), 2010 IEEE 26th International Conference on*, pages 485–496. IEEE, 2010.

[XYB+13] Wen Xiong, Zhibin Yu, Zhendong Bei, Juanjuan Zhao, Fan Zhang, Yubin Zou, Xue Bai, Ye Li, and Chengzhong Xu. A characterization of big data benchmarks. In *Big Data, 2013 IEEE International Conference on*, pages 118–125. IEEE, 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

TrendFashion -A Framework for the Identification of Fashion Trends

Samaneh Beheshti-Kashi[1], Michael Lütjen[2], Lennard Stoever[3], Klaus-Dieter Thoben[4]

Abstract: The fashion industry faces different challenges regarding accurate forecasts for future fashion products. The consumer demand is volatile and sales periods of fashion products are short due to production plants in Asia and target markets in Europe. Besides standard statistical approaches based on historical data and advanced methods such as the application of artificial neural networks or fuzzy logic, there are fashion experts, who use different information sources, e.g. fairs, social media, fashion websites, to predict design-trends as well as sales volumes. In this paper we follow this expert-driven approach by collecting data from fashion weblogs, news sites and fashion magazines, in order to identify actual and future design- trends. For this aim, we develop the TrendFashion Tool which collects data from these fashion sources and analyse them. On a higher level, this tool successfully separates fashion related posts from non-fashion related posts. And on a lower level, it identifies fashion related words and weights them according to an index.

Keywords: Fashion Forecasting, Social Media Analytics, Trend identification

1 Challenges in the Fashion Industry

The fashion industry has to handle diverse idiosyncrasies on different levels in order to avoid overstocked and stock-out inventories. External factors such as changing weather conditions, holidays or (sports) events have an impact on short-term customers' purchasing decisions and buying behaviours [Th10]. The availability of a product is crucial for fashion items, since buying decisions are often made at the point of sale [NGP13]. Moreover, different design trends are presented in fashion shows with a long-term range of approximately six months. Besides, street trends are emerging spontaneously with a mid-term range of approximately three months. Additionally, the most fashion products are shorter lived and will stay only for some weeks at the stores.

1 Universität Bremen, International Graduate School for Dynamics in Logistics (IGS), Hochschulring 20, 28359 Bremen bek@biba.uni-bremen.de

2 BIBA - Bremer Institut für Produktion und Logistik GmbH at the University of Bremen, Hochschulring 20, 28359 Bremen, Germany, ltj@biba.uni-bremen.de

3 minubo GmbH, Deichstraße 48–50, 20459 Hamburg, lennard@minubo.com

4 Faculty of Production Engineering, University of Bremen, Germany, tho@biba.uni-bremen.de

Product related factors such as the high variety in sizes and colours intensifies the demand uncertainty [CLP04]. In this regard, we can distinguish between traditional and fast fashion retailing. While traditional fashion retailing consists of producing one collection for the spring/summer and one for the autumn/winter season by producing the products in Asia and shipping them to Europe, the Spanish retailer Inditex is an excellent example of fast fashion retailing. Zara, the major division of Inditex has become the leader in rapid development of fast changing fashions by addressing latest street trends with products made in Spain with lead times less than several weeks. On the other hand, most production plants are placed in Asian countries where products can be produced cheaply [MTD11], but the time-to-market is long. Therefore, the reproduction of good selling products is rarely possible [Fs01]. One option for companies is to fly their goods to Europe or install additional production plants in Turkey or North Africa in order to be time efficient. However, the traditional as well as the fast fashion retailing strategy face challenges, which are worthy considering new approaches of trend identification and forecasting.

2 Related Work

2.1 Fashion Forecasting

The literature describes different approaches within the field of fashion forecasting. Due to the described characteristics of the fashion industry and the lack of historical data, standard forecasting methods such as simple statistical approaches often fail in providing accurate forecasting results. In recent approaches more advanced methods such as fuzzy logic or artificial neural networks are considered [SAC07] [Th10]. Further learning algorithm such as the extreme learning machine [Sz08] or the evolutionary neural network [ACY08] obtains accurate results. Though, in particular hybrid models achieve reasonable accuracy [WG10] [Ct14] [Tm14]. However, these works do not consider further influencing factors in their models. In contrast, Thomassey [Th10] provided a complex approach including external information such as weather conditions or information on promotions and events [Th10]. Considering these additional information, the accuracy of forecast results will increase, since the consumer demand is impacted by exactly these diverse factors. Another approach is followed by Mostard et al. [MTD11] and Teucke et al. [Tm14]. Both focus on pre-order information and include them into their forecasting models. Within the fashion industry the selection of the right colour is a highly important decision [Ki12]. Accordingly, this problem is also addressed by researchers. [Ct12] introduced a model for colour forecasting with very little data. In Gu et al. [GL10] a computer assisted colour forecasting database is presented. Beheshti-Kashi et al. [BT14] suggest integrating social media applications such as Twitter or weblogs into the fashion forecasting process [BT14]. This paper follows this approach and suggestion and introduces the so-called TrendFashion Framework.

2.2 Data Driven Approaches in the Fashion Industry

Within the last 5 years the fashion world has undergone a huge change. In particular, within the last 2-3 years new concepts such as outfittery[5], edited[6], AboutYou[7] or Zalon[8], Zalando's curated shopping solution, have emerged. All these applications require more consumer engagement, for instance, providing more personal data than for a traditional online purchase. Consequently, the companies obtain far more information about their customers than before. Information regarding each purchase, on colours, sizes, number of purchases, preferences, which products have been purchased together, or just product views are easily accessible. Consequently, a huge number of data is available about the consumer for the companies. The availability of these data, paves the way for data driven approaches within the fashion industry. In addition, to the available information provided by the consumer, further data is available throughout the internet. In recent times some start-ups[9] [10] have emerged who focus on these kinds of fashion related data.

Other approaches focus on a typical problem occurring in the fashion industry: sizes and the perfect fit. Every brand has its own particular sizes. Therefore, each garment or shoe ideally should be fitted before purchasing, which is not always possible, in particular in the case of online shopping. Recently, applications such as virtual try-on technologies or 3D Scanners have emerged which can be used by the consumers also at home in order to build their digital avatar. Both technologies are used for ensuring a more accurate fit for the consumer. In addition, to these applications augmented reality environment have obtain increased relevance. These environments allow the user for instance to browse a catalogue and try on apparels [BA15]. Ashdown and Loker consider the virtual try-on as a promising 3D technology for enhancing mass customization and online sales [AL10].

Google as the market leader in search engine technologies has conducted a study on fashion trends considering search queries. In the "Google Fashion Trends Report (U.S.)" they publish their results examining six billion apparel related queries in the United States. They categorize the extracted trends based on user search behaviour into the following six clusters applying Time Series Clustering: Sustained Growth, Seasonal Growth, Rising Stars, Sustained Decline, Seasonal Decline and Falling Stars [ZH15]. This work only considers search queries and not free and unstructured text. Therefore, no automated text processing and Natural Language Processing (NLP) methods are applied. Whereas, considering weblogs and digital fashion magazines require conducting NLP methods, in order to obtain valuable information.

[5] http://www.outfittery.de/
[6] http://www.edited.de/
[7] http://www.aboutyou.de/
[8] https://www.zalon.de/
[9] http://www.fashiongps.com/
[10] https://poshly.com/

2.3 Text Mining and Natural Language Processing

Text Mining methods and approaches have gained increased relevance within recent years. One reason is the huge emergence of free and unstructured data on the web. In particular, due to the numerous social media applications, which empowered every individual user to publish content, Text Mining is required in order to process the data and to extract valuable information. However, Text Mining methods are not limited to web usages, any textual database can profit from Text Mining. Text Mining includes several fields such as information retrieval, text analysis, information extraction, clustering, categorization, visualization, database technology, machine learning and data mining [TA99].

Granitzer [GR06] has identified the following four processes in Text Mining:

- Pre-processing: lexical analysis, part-of-speech tagging, stemming, stop word removal
- Information extraction: machine learning and linguistic analysis
- Feature generation: statistical analysis such as frequency analysis, word co-occurrences
- Operations on feature spaces: clustering, classification etc.

Within the last years the focus of Text Mining has gone towards Natural Language Processing (NLP). NLP is an older research field and involves technologies such as word stemming, lemmatization, multiword phrase grouping, synonym normalization and part-of-speech tagging [KP04].

3 The TrendFashion Framework

TrendFashion is a tool for automated collection and processing of fashion related web data. The current implementation focuses on German language weblogs, fashion magazines and news sites. Following the approach of identifying new design and street fashion trends by analysing web data, such as social media, news sites or digital fashion magazines different challenges have to be managed before the actual integration takes place. One essential task is to monitor the relevant sources. For instance, by continuously collecting and storing data from weblogs, we obtain an image of the fashion blogosphere within a certain time period. However, this data often includes unstructured text data and has to be pre-processed in order to obtain valuable output such as certain colours or fashion styles from it. Figure 3.1 illustrates the concept behind the tool. Within the database we have the following tables: websites, posts, single words and relation as well as the thesaurus. The websites table includes a list of URLs from weblogs, fashion magazines and news sites. With the help of RSS feeds and the crawling functionality we collect the data from these external sources and store them into the posts table. However, not all of the information is stored; we only consider the title, the published date, the URL and the actual content which has usually the format of free text (unstructured text).

Within the data processing phase, the content is analysed by the mean of NLP methods, and the paragraphs are separated into sentences. At this stage, we have included a "Moderelevanz" Index which weights the fashion relevancy of words and sentences. Relevant sentences are those which include for instance colours, brands or textile names. These sentences are separated into words. These words are matched with the thesaurus and a "Moderelevanz" Index is assigned to them. In the following sections, the collecting, processing as well as the thesaurus functionalities are presented.

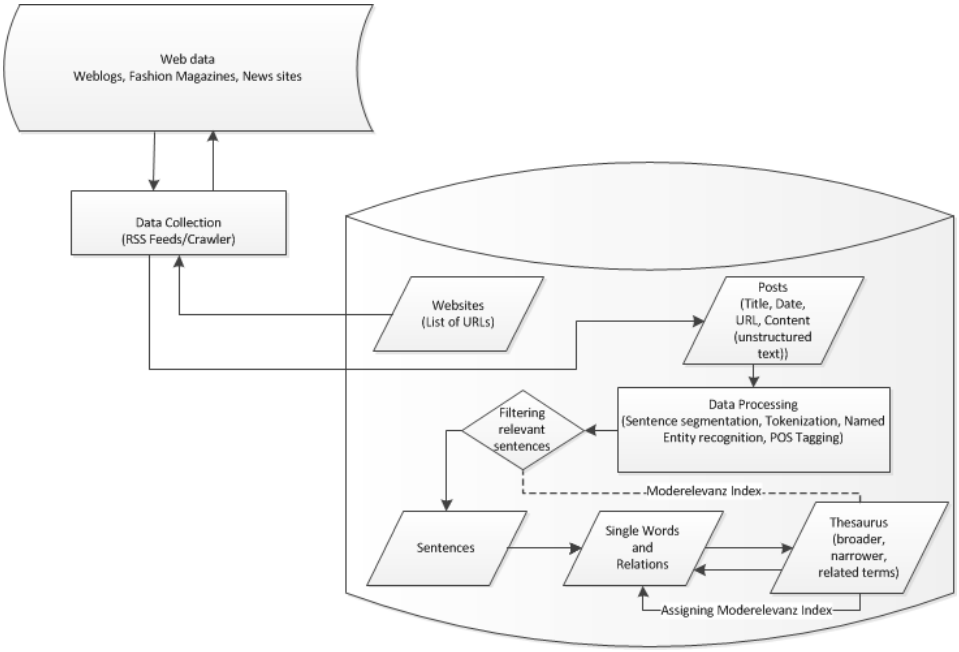


Fig. 3.1: Concept of TrendFashion

3.1 Collecting the Data

The tool is designed to handle different formats, namely RSS feeds and HTML-websites. The first approach was to use RSS feeds exclusively, since they always provide a complete time stamp. A detailed time reference to the written text is highly relevant for the given objective. The significance of the time aspect is due to the fast paced and rapidly changing fashion trends. Therefore, we have to focus on data from the same time period in order to correctly aggregate the data and to conclude the right information. Since a large number of the blogs do not provide RSS feeds, we added the second approach, HTML formats, to the tool. Currently, the tool considers 100 German

language weblogs, fashion magazines and news sites.

3.2 Processing the Data

After collection, the data has to be processed. For this task, we apply NLP methods such as sentence segmentation, tokenization, named entity recognition and Part-of-Speech-Tagging. The basis for further processing is the single posts which are stored as single documents. In the first step, the posts are separated into single sentences looking based punctuations. Then, the tokenization process will follow. After separating the sentences, the syntactic analysis starts. The Part-of-Speech-Tagger (POS-Tagger) annotates of every single word of sentence to a specific part-of-speech such as noun, verb or adjective. For the Tagging process we selected the German version of Wiktionary.org, which is a free online wiki-based dictionary, for identifying the correct part-of-speech. The "Moderelevanz" Index, filters and weight the separated POS according to their relevance to fashion topics and filter out stop words such as articles or pronouns at the same time. The extracted and tagged words are stored in a separated table.

3.3 Thesaurus

In order to enhance the semantic functionality of the tool, we have included a thesaurus. Such a thesaurus is often designed for a particular domain, and in contrast to a dictionary it provides relations between words. These relations are required for automated text analysis and are based on several taxonomies in which we classified different product groups. So far, we have included taxonomies only on a high level. Detailed taxonomies are in progress and will be integrated for enhancing the learning capability of the tool. The thesaurus includes broader, narrower and related terms. In addition, to these corresponding terms, the words are stored according to their "Moderelevanz" Index. Furthermore, the part of speech of each entry is provided. In the current version of the tool, the "Moderelevanz" Index is adjusted and adapted manually.

4 Results

The TrendFashion Database handles different tasks. It collects and processes the data. It filters the content from the noise of a website (navigation, HTML tags etc.) and stores each post separately. Based on the posts, sentences are separated; from these sentences words and word combinations are extracted and stored. Besides the collection and storage of posts, sentences, and words, it provides lists of word occurrences. This is the first step in order to extract future trends. Though, a further focus is on extracting words belonging together, in order to increase the semantic ability of the tool. For identifying fashion trends, adjectives such as colours are crucial. Therefore, it is important to find adjective-noun or adjective-adjective-noun combinations belonging together. In the current version, the tool is able to identify and extract such combinations in a reasonable

way. Figure 4.1 illustrates matched word combinations in the area of evening robe, glitter and sparkling clothing. The font size of the word combinations correspondents to their frequency rates: the larger the font size, the higher is their occurrences within the text. In the most cases, the tool matched adjective-noun combinations. However, in the example of *femininenSmokingLook,* we can see that in a first step the two nouns *Smoking* and *Look* are combined, and in the second step the adjective *feminine* is matched to the new combination.



Fig.4.1 Extracted word combinations

With the word cloud in Figure 4.1 a broader and more general view on the extracted word combinations is presented. However, the tool can also identify more specific combinations and relations. Since in fashion just a slight variation of a certain colour, shape or cut will have an impact on the success of a product, we assume that for the prediction of a future design trend more specific features on the colour, cut or style are required and not just a general overview of a product. For instance, looking at the product "blouse", the following adjective noun combinations are extracted *bodenlange Bluse*, *rosafarbene Bluse*, *bieder Bluse, hochgeschlossene Bluse.* In this example, different features describing a "blouse" are provided: a length (*bodenlange*), a colour (*rosafarbene*), a shape (*hochgeschlossene*) as well as information on the style (*bieder*). Similarly, the tool extracts diverse adjective noun combinations for the product jeans. We consider the presented tool as the basis for future predictions. Since the objective is

to provide trend information also on a detail level, it is crucial to classify the extracted information according to the different features. This will be also done by the help of taxonomies. Figure 4.2 illustrates the extracted word combinations as a hierarchal tree. In this case, we have taken the features colour, cut, style as well as additional describing attributes and matched the corresponding adjectives to them.

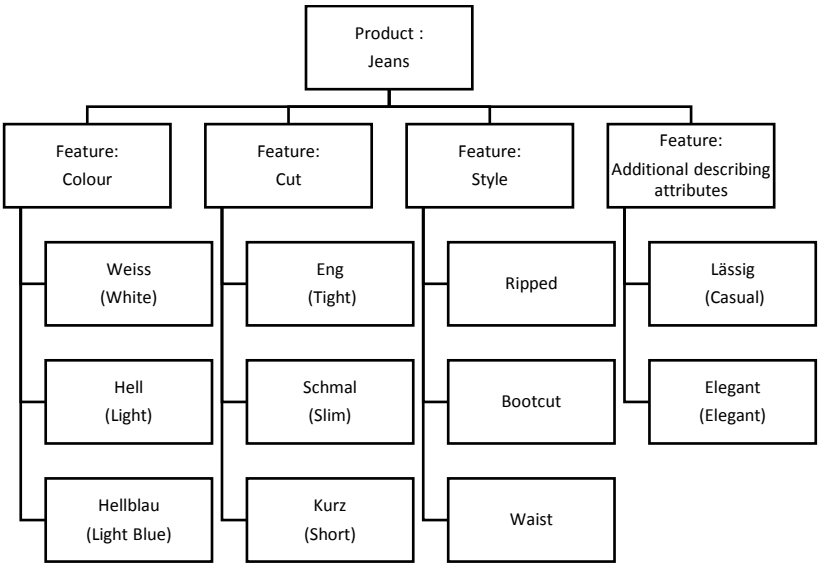


Fig. 4.2: Extracted features on the product jeans

Furthermore, we examined 22 webpages about information on future design trends for the autumn winter season 2015/2016. Figure 4.3 illustrates the identified word combinations extracted from the text provided on the 22 webpages. In contrast to figure 4.2, it shows that more detailed information on a certain product can be extracted. We can find information on the material (*FellPelz*), on certain components of a shoe such as the sole (*dickeSohlen*) or on the heel (*hochAbsatz* or *flacheSchuh*).



Fig.: 4.3: Extracted word combinations on shoe features

5 Conclusion and Future Work

In this paper we propose a data driven approach to predict future design trends. The presented tool handles collecting and processing free and unstructured text data from the web. For this paper, we consider German language weblogs, news sites and digital fashion magazines. The data is collected, stored and analysed by the mean of NLP methods. Extracting and matching of adjective noun combinations are conducted automatically. With the included "Moderelevanz" Index fashion related content can be filtered from the noise to a certain level and is currently manually adapted. The automatic assignment of the Index is considered as the next step. Furthermore, we fill the thesaurus with more detailed taxonomies in order to enhance the semantic functionality. In addition, the time dimension will be added to the analysis in order to be able to build up the trends and to follow them. In the next step, the number of input websites will be increased, in order to work on a larger sample and consequently covering potential niches. Further future work additionally includes extending to English web data, and also to focus on a multi-language approach.

References

[ACY08] Au, K.-F., Choi, T.-M., & Yu, Y. : Fashion retail forecasting by evolutionary neural networks. International Journal of Production Economics, 114(2), 615–630, 2008.

[AL10] Ashdown, S., Loker, S.; Mass-Customized Target Market Sizing: Extending the Sizing Paradigm for Improved Apparel Fit. Fashion Practise 2/2, 147-174.

[BA15] Baytar, F., Ashdown, S.: An Exploratory Study of Interaction Patterns around the Use of Virtual Apparel Design and Try-on Technology. Fashion Practise: The Journal of Design, Creative Process and the Fashion Industry, 7/1, 31-52, 2015.

[BT14] Beheshti-Kashi, S., Thoben, K.-D.: The usage of social media text data for the demand forecasting in the fashion industry fashion sales forecasting and the predictive power of online. In Dynamics in logistics. Fourth international conference, LDIC 2014 Bremen, Germany, February 2014 Proceedings, 1–4, 2014.

[CLP04] Christopher, M., Lowson, R., Peck, H.: Creating agile supply chains in the fashion industry. International Journal of Retail & Distribution Management, 32/8, 367–376, 2004.

[Ct14] Choi, T., Hui, C., Liu, N., Ng, S., & Yu, Y.: Fast fashion sales forecasting with limited data and time. Decision Support Systems, 59, 84–92, 2014.

[Ct12] Choi, T., Hui, C., Ng, S., & Yu, Y.: Color Trend Forecasting of Fashionable Products with Very Few Historical Data, 42/6, 1003–1010, 2012.

[Fs01] Fissahn, C.: Marktorientierte Beschaffung in der Bekleidungsindustrie - Schriften zur Textilwirtschaft, Band 53,. FATM, Münster, 2001.

[Gr 06] Granitzer, M.: KnowMiner: Konzeption und Entwicklung eines generischen Wissenserschliessungsframeworks, PhD Thesis TU Graz, Graz, 2006.

[GL10] Gu, W., & Liu, X.: Computer-Assisted Color Database For Trend Forecasting, 0–3, 2010.

[Ki12] King, J. A.: Colour trends and selection in fashion design. Colour design: Theories and applications, 535–550, Woodhead Publishing Limited, 2012.

[KP04] Kao, A., Poteet, S.: Text Mining and Natural Language Processing – Introduction for the Special Issue. SIGKDD 7/1, 2004.

[MTD11] Mostard, J., Teunter, R., & de Koster, R.: Forecasting demand for single-period products: A case study in the apparel industry. European Journal of Operational Research, 211/1, 139–147, 2011.

[NGP13] Nenni, E. M., Giustiniano, L., & Pirolo, L..: Demand Forecasting in the Fashion Industry: A Review. International Journal of Engineering Business Management, 5, 2013.

[SAC07] Sun, Z.-L., Au, K.-F., & Choi, T.-M. : A neuro-fuzzy inference system through integration of fuzzy logic and extreme learning machines. In: IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics : A Publication of the IEEE Systems, Man, and Cybernetics Society, 37/5, 1321–31, 2007.

[Sz08] Sun, Z.-L., Choi, T.-M., Au, K.-F., & Yu, Y.: Sales forecasting using extreme learning machine with applications in fashion retailing. Decision Support Systems, 46/1, 411–419, 2008.

[TA99] Tan, A.-H.:Text Mining: The state of the art and the challenges, PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases, Beijing, (1999), 65-70, 1999.

[Th10] Thomassey, S.: Sales forecasts in clothing industry: The key success factor of the supply chain management. International Journal of Production Economics, 128/2, 470–483, 2010.

[Th14] Thomassey, S.: Sales Forecasting in Apparel and Fashion Industry: A Review. In. (Choi, T.M. et al.): Intelligent Fashion Forecasting Systems: Models and Applications. Springer-Verlag Berlin Heidelberg 2014, 9-27, 2014.

[Tm14] Teucke, M., Ait-alla, A., El-berishy, N., Beheshti-Kashi, S., & Lütjen, M.: . Forecasting of Seasonal Apparel Products. In (H. Kotzab, J. Pannek, & K.-D. Thoben): Dynamics in Logistics. Fourth International Conference, LDIC 2014 Bremen, Germany, February 2014 Proceedings. Springer, 2014.

[WG10] Wong, W. K., & Guo, Z. X.: A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. International Journal of Production Economics, 128/2, 614–624, 2010.

[ZH15] Zimmer, O., Horwitz, Y.: Google Fashion Trends Report (U.S.), 2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Dynamic Domain-Specific Language for BigData Tasks' Description

Artem Zakharchuk[1], Sergey Kovalchuk[2], Denis Nasonov[3]

Abstract: The paper presents a dynamic Domain-Specific Language (DSL) which is developed to provide the capability of high-level BigData task descriptions within e-Science applications. The dynamic structure of the DSL supports language structure extension depending on a particular problem domain defining specific requirements, data processing, and aggregation and simulation procedures. The extension is implemented using the set of domain-specific libraries providing integration into the DSL interpretation system. The developed DSL is integrated with cloud computing environment CLAVIRE and DSLs' family used to describe composite application, software packages and resources within this environment. Together with knowledge-based technologies implemented on the basis of this environment (Virtual Simulation Objects - VSO, knowledge-based languages, etc.) the DSL is used to develop and execute workflows within a distributed cloud environment combining computational-intensive tasks with data-intensive MapReduce processing within common semantic space defined by domain knowledge.

Keywords: domain-specific language, big data, map-reduce, cloud computing, code-to-data

1 Introduction

Contemporary e-Science tasks often require data-intensive applications to be incorporated as a major part of the composite solution [He09]. Today a huge amount of various data (stored in the data warehouses, accessible as observation of forecasting services, real-time monitoring data, etc.) is available for scientific analysis, information retrieval and knowledge discovery. The growth of volume and variety of available data together with processing performance requirements causes the emergence of the BigData area [PZ14]. The BigData area collects a set of practices, methods, technologies which are aimed to manage better the large data arrays. For example, the MapReduce model [DN14] is considered today as a standard solution for distributed processing of large data arrays. Nevertheless BigData still has a lot of issues to be solved. One of the questions currently discovered within the area of BigData is high-level user-friendly task definition [As14]. A common way of high-level task description is Domain-Specific Languages (DSL) like Pig (https://pig.apache.org/) or Hive (https://hive.apache.org/). Moreover the DSL can serve as a core technology for solving data and operation fusion problem combining high-level task description with performance optimization (see e.g. [Ac12]). Nevertheless complex e-Science tasks often require high-level integration of diverse data

[1] ITMO University, Saint-Petersburg, Russia, zakharchukart@gmail.com
[2] ITMO University, Saint-Petersburg, Russia, kovalchuk@mail.ifmo.ru
[3] ITMO University, Saint-Petersburg, Russia, denis.nasonov@gmail.com

and software (often available within a distributed environment) which makes a high-level task description more complicated. Within the e-Science area the common approach to description of the complex domain-specific tasks is workflow (WF) [Gi07]. Still the WF concepts usually exploit the data-to-code approach (data is transferred to the executable services) when the BigData solutions are usually based on the code-to-data concept (executable modules are transferred to the data sites) [Ma10]. The differences in the approach make it difficult to integrate seamlessly the BigData solutions into the WF management systems (WFMS): the existing solutions usually have a strong conceptual border between these two conceptual approaches, e.g. by considering the BigData subsystem as a special kind of service integrated into WF [BBB13]. On the other hand, nowadays e-Science solutions tend to deviate from pure WF concept (as mostly a programming concept) and shift towards the approaches more friendly to end-users (mostly domain scientists) who often have a lack of programming background [Mc09]. Moreover today's trends require the e-Science solution to be enhanced from procedural WF description towards complex system-level investigation [FK06].

Within the presented work we try to develop a solution which enables the end-user to define high-level tasks which combine seamlessly both data-to-code and code-to-data approaches by the use of knowledge-based support for simulation-based scientific exploration [SKB13]. As a core part of the solution we have developed a dynamic domain-specific language (DSL) which can adapt to the particular problem domain and might be integrated into the cloud computing environment to support the BigData tasks' definition. The paper presents the main features of the developed language and its implementation for composite application development.

2 Solution Development

2.1 Code and Data Relationship

The code-to-data and data-to-code approaches consider two artifacts usually involved into application development. Within our approach each of these artifacts can be either provided by the user (considered as portable) or provided by a reference to remote site (abstract definition possible). This permits us to identify four situation and popular solutions related to each (see Table 1).

The presented features and systems are mentioned to highlight diversity of the approaches which are used within different e-Science applications. Within the presented work we consider a task as a high-level description of the problem which contains or can be interpreted into provided data (PD) and pieces of code (PC). To solve the task, the distributed environment, which contains computational services (LC) and data stored within distributed storage or available on demand using remote services, were used.

	Provided code (PC)	Linked code (LC)
Provided data (PD)	Simple application. Code and data are portable.	Data-to-code. Regular WFMS. Abstract services' composition. Distributed environments (grid, cloud).
Linked data (LD)	Code-to-data. BigData analytics' solutions. Distributed data storage with execution facilities, data services.	High-throughput solutions. Streaming-based technologies for data processing.

Tab. 1: Code and data processing

The distributed data, available for processing can be considered as informational resource provided to the user on demand (where the request is formalized within the provided code). This point of view considers the processing of BigData within the scope of cloud computing [Fo08] approach. Mapping the existing cloud computing Maturity Models (CCMM) (see e.g. [Gtsi]) which describes the level of cloud computing concept's implementation onto the data processing task can define several stages of data provision: ***Consolidation.*** Having a lot of *data sources*, a large volume of data storage could be formed and updated using external services, public data warehouses, simulation archives, etc; ***Abstraction.*** Unified access to the data with different formats can be provided by solution which supports *abstract data types* by automatic mapping of data requests onto the available data stored or provided as a service; ***Automation.*** Use of the MapReduce model requires from the user to develop an executable module. As the typical user of e-Science solution usually has a lack of technological background, the automation level should provide *high-level expressive tools* for task definition; ***Utility.*** This level should support the user with extended *analytics toolbox* statistical aggregation, data mining, etc; ***Cloud.*** Finally the data become the subject to operate in a market way where the cloud system automatically processes the request from the user providing the results of the task solving and hiding all the low-level procedures.

The CCMM levels can define the key waypoints during the development of solution architecture which can serve as a high-level e-Science tool to efficiently support the user during the scientific investigation.

2.2 BigData within e-Science Tasks

Considering the typical e-Science tasks, BigData requirements appeared within such tasks, and CCMM level mentioned earlier, the following features can be defined for the developed solution:

- *High-level semantic description* of available data to integrate various data sources and to provide the user with capability to work with abstract data types defining them with domain-specific semantics.
- The semantic description should link the data with the *structure of system* being analyzed and thus support the system-level science investigation.
- Generally the e-Science tasks are solved within Simulation-Driven Approach (SDA). The data processing should be *integrated with the simulation procedures* a) by being incorporated into the hybrid high-level composite application (inferred from the high-level task definition); b) by performing simulation procedures in code-to-data way within the data storage.
- *Domain-specific languages* (DSL) [DKV00] provide a powerful toolbox for high-level interaction with the user. Still usually the domain-specific language is developed for the specific problem.
- Considering the e-Science area a huge variety of domains and tasks can be identified. Thus the developed DSL should be *dynamically adoptable* to the particular problem domain and task.

2.3 Solution's Architecture

The proposed solution is based on the idea of knowledge-based expressive toolbox [Ko13] where the hierarchy of languages (with textual or graphical notation) is defined by the set of knowledge from different problem domains. The basic hierarchy of expressive tools includes five levels: L1 – services description; L2 – services composition (WF); L3 – simulation objects description; L4 – objects composition (description of investigated system; L5 – high-level task definition.

To extend this architecture the parallel tier (see fig. 1) is presented which describes the code-to-data part of composite application built during high-level task interpretation. To implement this tier the dynamic DSL is developed which serves to describe the high-level BigData tasks. It incorporates the imperative procedures, defined by the task being solved and applies them to the structure of investigated system. As the structure is linked with the available data, the interpretation of the language can automatically generate the full executable MapReduce code which is ready to run on the data storage system. The code-to-data part of the composite application is interconnected with the data-to-code part in several ways: a) the system's structure is shared between these two parts and used both to generate general-purpose executable WF and to select the proper data during data analysis; b) dynamic dataset, which is updated during simulation and data analysis, is sharable which enables to join the parts and apply them separately; c) distributed data storage can be used to transfer small datasets to the computational services (in data-to-code way); d) finally the cloud management system can control the deploy of portable software to the data storage nodes to support short local runs for the simulation purposes (see the requirements).

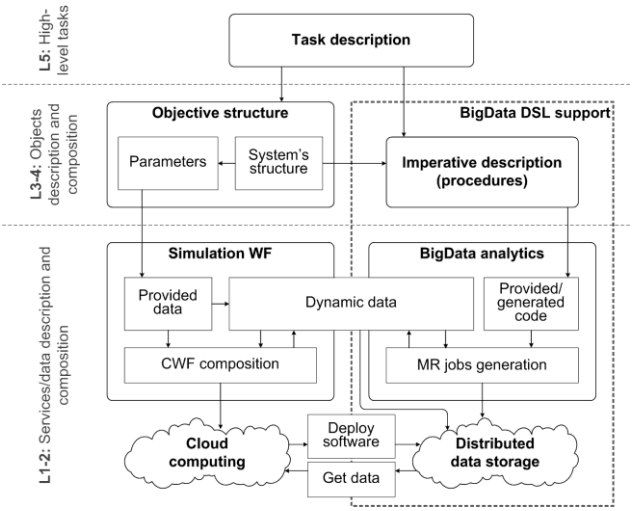


Fig. 1: Hybrid composite application

The dynamic DSL being developed within the knowledge-based approach incorporates the following artifacts:

- ***Domain-specific semantics.*** A set of domain-specific objects which are used to describe the structure of investigated system is defined. This part of knowledge enables interconnection with WF and simulation procedures on the levels L4-5.
- ***Data formats.*** To support the integration and high-level definition of the task, the semantically marked data of format description and atomic analysis procedures should be described.
- ***Data aggregation patterns.*** To interpret imperative description of procedures within the high-level task with MapReduce code generation, the data aggregation and statistical analysis procedures should be available.
- ***Cloud infrastructure knowledge*** should be provided to support a) further processing of the results by the other parts of composite application; b) call the local software described in the same way as regular cloud computing services.

3 Implementation details

The developed solution is based on CLAVIRE [Kn12] cloud computing environment which actively used knowledge-based technologies for solving-Science tasks. It supports a set of DSLs for description of available software services (EasyPackage language based on Ruby) and develops high-level composite applications in form of abstract WF (EasyFlow declarative language). Additionally the Virtual Simulation Objects (VSO)

technology was developed to support high-level semantic description of investigated system which can be mapped on AWF structure and executed within the cloud environment [Ko12]. The platform uses distributed data storage (DStorage) which was extended to support MapReduce operations (described as Java code) on the stored datasets [KRS13]. Within the presented research we developed a dynamic BigData DSL which serves as an expressive tool for definition of high-level data analytics tasks by the use of the mentioned technologies. The goal of the language is to extend the implemented CLAVIRE-based hierarchy of languages to support the development of hybrid composite applications exploiting both data-to-code and code-to-data approaches. The DSL was developed on the basis of ANTLR which is cooperated with Java environment and provides powerful features for DSL development. The following subsection describes the developed solution with more details.

3.1 DSL Structure

Current version of the dynamic DSL is being developed for CLAVIRE platform, in particular for execution within distributed storage (DStorage). The basic architecture of the distributed data processing within CLAVIRE and DStorage is presented in the fig. 2.

The data processing solution implements fixed MapReduce scheme which allow us to process data files stored on the nodes of the distributed storage. The data processing module is implemented by Java programming language and uploaded to the storage with automatic replication onto the whole set of storage nodes. In order to run data processing task a request has to be processed by the core of data storage which accepts configuration and additional parameters (query strings, additional data etc.) and runs distributed processing of files with aggregation of the results. The distributed data processing solution is used as a basic technology which is extended by the means of DSL integration. Within the storage architecture the DSL script is processed as a parameter of distributed task while the DSL interpreter is implemented as a data processing module and uploaded to the storage in a general way.

In order to support semantic integration and high-level concepts' semantics the VSO technology is used as one of the extensions of the CLAVIRE platform. The VSO concept and technology provide the high-level semantic abstractions which integrate domain-specific objective description of investigated system and simulation infrastructure. The integration with CLAVIRE enables the use of the abstract WFs in a form of extended EasyFlow (DSL for WF description) (grammar is shown in fig. 3) and high-level domain-specific parameters of available software packages described using EasyPackage (DSL for abstract software services description).

In MapReduce scheme a map procedure defines processing of files within the storage, while reduce procedure defines aggregation of the mapping results. As the data is distributed, the analysis functions may be applied to them only when certain conditions are fulfilled. These conditions should be described in the domain library, which is dynamically linked to DSL.

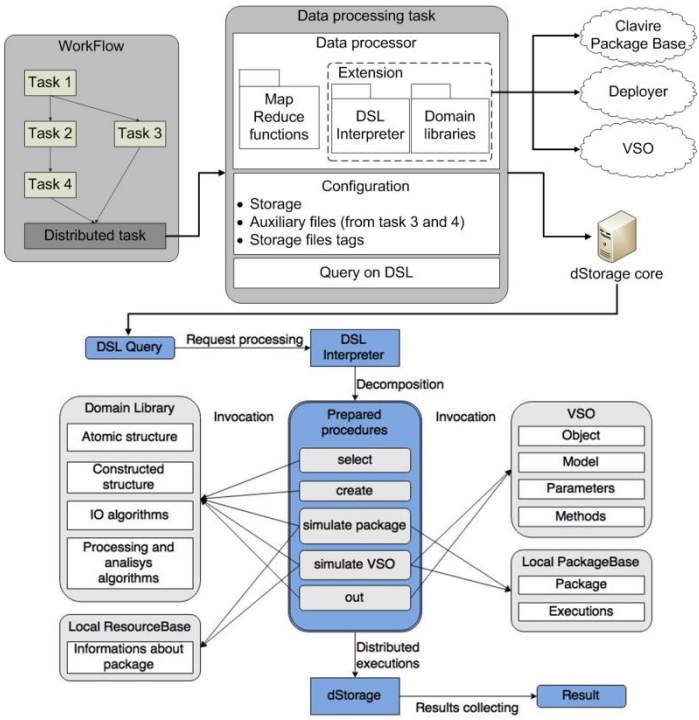


Fig. 2: Distributed data processing within CLAVIRE platform and query interpretation

```
Application ::= {InputData} {WFAttributes} {({Comment}|{BlockDesc})}
Comment ::= "//" String
InputData ::= require IdsList ";"
WFAttributes ::= "{" flow ":" Id "=" Expression "}"
BlockDesc ::= {Attribute} (StepBlock|ScriptBlock) [PostProcessing] ";"
StepBlock ::= (step|~step)Id runs CompositeId [after IdsList] ParametersList
ScriptBlock ::= script Id runs Id [after IdsList] ParametersList ScriptSection
ScriptSection ::= "{:" Code ":}"
CompositeId ::= {Id [Point]}
IdsList ::= {Id [ "," ]}
Attribute ::= "[" Id "=" Expression "]"
ParametersList ::= "(" {(Parameter|Port)[,]} ")"
Parameter ::= Id "=" Expression
Port ::= Id "< -" ObjectField
PostProcesing ::= post code ruby RubyCode code end
Expression ::= Literal | ObjectField
ObjectField ::= CompositeId ["[" Expression"]"]
```

Fig. 3: Simplified grammar of extended EasyFlow

Exactly these libraries determine the dynamic of the DSL. The domain library should contain the following items:

- The atomic structure of data stored in files.
- Constructed data structures are generated during data analysis.
- Unified algorithms for reading data structures from the storage

- Algorithms for processing and analysis of data structures (1, 2).

Also, during execution of data processing, the various software packages should be able to run. It is constructed by using EasyPackage description within local PackageBase and local ResourceBase of CLAVIRE platform.

Simplified grammar of developed dynamic DSL is shown in figure 4.

```
Application ::= {InputData} Select [Create] [Simulate] Out
InputData ::= Id " = " value
Select ::= "select" Id "where {" {Condition} "}"
Create ::= "create" Id "from" Id "where {" Condition "}"
Simulate ::= "simulate" ("VSO"|"Package") Id "from" Id ["[" IdList "]"] ParametersList
Out ::= "out" outParam {"," outParam}
Condition ::= Id sign Expression
IdList ::= {Id}
ParametersList ::= Id " = " Expression
OutParam ::= Id "as" Id
```

Fig. 4: Simplified grammar of developed DSL

3.2 Query Interpretation

Query written with the DSL and described with a specific problem analysis and processing data, is transmitted together with other parameters at startup of data processing task. This script should consist of keyword of the DSL and domain-specific keywords. Procedures Map and Reduce for distributional data processing is generated under the condition of this request. Processing the DSL query is shown in fig. 2.

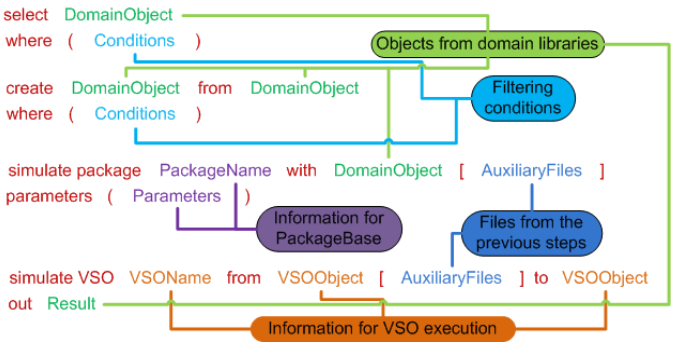


Fig. 5: Skeleton of DSL grammar

Currently the DSL contains the following keywords: ***Select.*** Creates instances of the objects, which are described in the domain library and fill with data from storage using reading algorithms; ***Where.*** Analyzes the object using algorithms from the domain library; ***Create.*** Creates the object from other (objects); ***Simulate.*** Executes software packages or VSO by using Local PackageBase and Local ResourceBase; ***Out.*** Presents the results.

In general, the skeleton of query script is shown in fig. 5. Here beside the basic DSL keywords the following structures are presented for the purpose of dynamic extension of the language structure: a) references to semantically defined objects, parameters and models are provided using VSO technology or software packages for local calls defined using EasyPackage within the CLAVIRE environment; b) objects, procedures and conditions provided within domain libraries; c) data file semantic identification which identify the datasets within DStorage.

4 Hybrid Workflow Case Study

Experimental studies in cyclone's analysis have significant importance for protection of Saint-Petersburg area from storm surges [AK07]. One of the vivid examples of such importance is the hybrid composite application that is described in the Listing 1.

Listing 1. Hybrid composite application for cyclones' analysis

```
require BSHFile, projects, inControlPoints, points, searchedhirlam;
script CycloneSearch runs BDDSL (
  dir = "/data/path", tag = "ERA"
)
{:
  select Cyclone where (
        trajectory_path = "(65.67, 8.19), (65.97, 23.45), (60.97, 31.45)"
        accuracy = "(5, 5)"
  )
  out Cyclone.Parameters() as outs
:};
step CycloneStatistialAnalysis runs scilab (
  input_folder = CycloneSearch.Result.outs["data/out"],
  script_name = "SCI_stat.sce"
);
script FloodSimulation runs BDDSL (
  dir = "/data/path", tag = "HIRLAM",
  config = CycloneStatistialAnalysis.Result.outs["output.dat"],
  points = points, projects = projects, BSHFile = BSHFile,
  searchedpattern = searchedhirlam, inControlPoints = inControlPoints
)
{:
  select Forecast where (
        preasure field = searchedpattern
  )
  simulate package cyclonegenerator with parameters (
    inConfigFile = configFile
    forecastField = Forecast.field
  )
  simulate VSO FloodSim from cyclonegenerator.Result [points, projects,
           BSHFile, inControlPoints] to FloodSimulation
  out FloodSim.Result as out
:};
step FloodStatistialAnalysis runs scilab
(
  input folder = FloodSimulaiton.Result.outs["/out/"],
  script_name = "FLOOD_stat.sce"
);
```

During the first step all cyclones that match to the provided criteria on trajectory path and its accuracy are detected. In the composite application this task is executed in the step CycloneSearch, that, firstly, extracts all the cyclones that are stored in ERA (http://apps.ecmwf.int/datasets/data/era40-daily/) files (operation "select"). Then a filtration (operation "where") of the found cyclones occurs, using parameters of trajectory path and accuracy that are extracted with the function Parameters(). All these actions are performed by using domain library Cyclone on the distributed storage nodes.

Statistically estimated parameters for the most probable cyclones' emersion are calculated at the second step. For this purpose the step CycloneStatisticalAnalysis evaluates scilab package (software for numerical computation) in the general way (data-to-code).

During the third step all forecasts that are similar to the one provided by the user are detected and the cyclone with the estimated parameter is simulated. The step FloodSimulation is also executed in the distributed storage. Firstly, the forecasts, which comply with the conditions on the provided pressure (file "serchedhirlam"), are selected from HIRLAM (http://www.hirlam.org/) data. This part of the step uses the domain library Forecast. Secondly, package cyclonegenerator is executed locally on the storage agent, with the parameters from "configFile" file that was obtained in the previous step CycloneStatisticalAnalysis. The result of this package is transmitted to the VSO SeaForecast with other input data. Due to VSO two packages SWAN (wave model, http://swanmodel.sourceforge.net/) and BSM (Baltic Sea Model, generate water level forecast) are running subsequently on the storage agent. Finally water level forecasts are returned to the step FloodStatistialAnalysis for analysis.

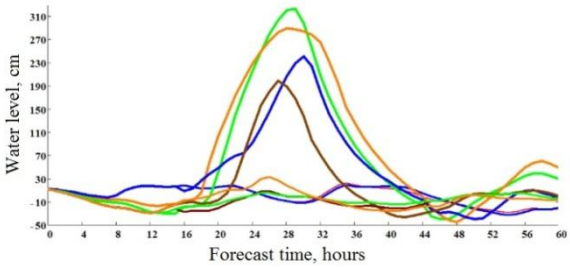


Fig. 6: Forecasted water level in Saint-Petersburg for generated cyclones

On the last step of the analysis of all water level forecasts is performed in FloodSimulation. This package is executed in the general way (data-to-code).

The results of completed composite application are presented on figure 6 (the plot was generated automatically by the last step of the application). In spite of the fact that on all found forecasts the cyclone with the same parameters was generated, the form and the height of the expected flood pick is different. It can be explained by two main reasons: pressure field influence by itself (part without cyclone) and simplicity of cyclone

generation algorithm (it doesn't take into account).

5 Conclusion

The dynamic DSL being developed provides a flexible expressive tool for high-level definition of data analysis tasks. The knowledge-based technologies used to interpret the DSL structure enables the interpretation of the terms and structures defined by the semantics of particular problem domain (or multiple domains in case of multidisciplinary tasks). The DSL is integrated with cloud computing environment CLAVIRE and enables the development of hybrid solutions by integrating the DSL with languages for WF definition. The complex system for high-level definition of the tasks allowing the automatic building of hybrid composite applications is now being developed on the basis of the presented DSL technology integrated with the existing expressive technologies of CLAVIRE platform.

Acknowledgements. This paper is financially supported by Ministry of Education and Science of the Russian Federation, agreement #14.578.21.0077 (24.11.2014). Data management facilities were developed within the project "Big data management for computationally intensive applications" (project #14613).

References

[As14] ASSUNÇÃO, MARCOS D. ; CALHEIROS, RODRIGO N. ; BIANCHI, SILVIA ; NETTO, MARCO A.S. ; BUYYA, RAJKUMAR: Big Data computing and clouds: Trends and future directions. In: *Journal of Parallel and Distributed Computing* (2014)

[Ac12] ACKERMANN, S ; JOVANOVIC, V ; ROMPF, T ; ODERSKY, M: Jet: An Embedded DSL for High Performance Big Data Processing. In: *International Workshop on End-to-end Management of Big Data (BigData 2012)*

[AK07] AVERKIEV, A. S. ; KLEVANNY, K. A.: Determining cyclone trajectories and velocities leading to extreme sea level rises in the gulf of Finland. In: *Russian Meteorology and Hydrology* Bd. 32 (2007), Nr. 8, S. 514–519 — ISBN 1068373907080

[BBB13] BARANOWSKI, MIKOLAJ ; BELLOUM, ADAM ; BUBAK, MARIAN: Mapreduce operations with WS-VLAM workflow management system. In: *Procedia Computer Science*. Bd. 18, 2013, S. 2599–2602

[DKV00] DEURSEN, ARIE VAN ; KLINT, PAUL ; VISSER, JOOST: Domain-specific languages: an annotated bibliography. In: *ACM Sigplan Notices* Bd. 35 (2000), Nr. 6, S. 26–36 — ISBN 0362-1340

[DN14] DOULKERIDIS, CHRISTOS ; NØRVÅG, KJETIL: A survey of large-scale analytical query processing in MapReduce. In: *VLDB Journal* Bd. 23 (2014), Nr. 3

[FK06] FOSTER, I. ; KESSELMAN, C.: Scaling System-Level Science: Scientific Exploration and IT Implications. In: *Computer* Bd. 39 (2006), Nr. 11 — ISBN 0018-9162

[Fo08] FOSTER, I ; ZHAO, YONG ; RAICU, I ; LU, S: Cloud Computing and Grid Computing 360-Degree Compared. In: *Grid Computing Environments Workshop, 2008. GCE '08*, 2008 — ISBN VO -, S. 1–10

[Gi07] GIL, YOLANDA ; DEELMAN, EWA ; ELLISMAN, MARK ; FAHRINGER, THOMAS ; FOX, GEOFFREY ; GANNON, DENNIS ; GOBLE, CAROLE ; LIVNY, MIRON ; U. A.: Examining the challenges of scientific workflows. In: *Computer* Bd. 40 (2007), Nr. 12, S. 24–32 — ISBN 0018-9162 VO - 40

[Gtsi] GTSI.: *GTSI Cloud Computing Maturity Model*. URL http://www.gtsi.com/cms/documents/White-Papers/Cloud-Computing.pdf

[HTT09] HEY, TONY ; TANSLEY, STEWART ; TOLLE, KRISTIN: The Fourth Paradigm: Data-Intesive Scientific Discovery Bd. 99 (2009), Nr. 8 — ISBN 9780982544204

[Kn12] KNYAZKOV, KONSTANTIN V. ; KOVALCHUK, SERGEY V. ; TCHUROV, TIMOFEY N. ; MARYIN, SERGEY V. ; BOUKHANOVSKY, ALEXANDER V.: CLAVIRE: e-Science infrastructure for data-driven computing. In: *Journal of Computational Science* Bd. 3 (2012), Nr. 6, S. 504–510

[KRS13] KOVALCHUK, S. V. ; RAZUMOVSKIY, A. V. ; SPIVAK, A. I.: Domain-specific technology for big data storing and processing within cloud computing platform CLAVIRE. In: *Dinamika Slozhnyh Sistem (in Russian)* Bd. 3 (2013), S. 106–109

[Ko12] KOVALCHUK, SERGEY V. ; SMIRNOV, PAVEL A. ; KOSUKHIN, SERGEY S. ; BOUKHANOVSKY, ALEXANDER V.: Virtual simulation objects concept as a framework for system-level simulation. In: *2012 IEEE 8th International Conference on E-Science, e-Science 2012*, 2012 — ISBN 9781467344678

[Ko13] KOVALCHUK, SERGEY V ; SMIRNOV, PAVEL A ; KNYAZKOV, KONSTANTIN V ; ZAGARSKIKH, ALEXANDER S ; BOUKHANOVSKY, ALEXANDER V: Knowledge-based Expressive Technologies within Cloud Computing Environments. In: *8th International Conference on Intelligent Systems and Knowledge Engineering (ISKE2013)*, 2013, S. 1–10

[Mc09] MCPHILLIPS, TIMOTHY ; BOWERS, SHAWN ; ZINN, DANIEL ; LUDÄSCHER, BERTRAM: Scientific workflow design for mere mortals. In: *Future Generation Computer Systems* Bd. 25 (2009), Nr. 5, S. 541–551 — ISBN 9781605587172

[Ma10] MANJUNATHA, ASHWIN ; RANABAHU, AJITH ; ANDERSON, PAUL ; SHETH, AMIT: Getting Code Near the Data: A Study of Generating Customized Data Intensive Scientific Workflows with Domain Specific Language. In: *2nd IEEE International Conference on Cloud Computing Technology and Science* (2010)

[PZ14] PHILIP CHEN, C. L. ; ZHANG, CHUN YANG: Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. In: *Information Sciences* Bd. 275 (2014), S. 314–347 — ISBN 0020-0255

[SKB13] SMIRNOV, PAVEL A. ; KOVALCHUK, SERGEY V. ; BOUKHANOVSKY, ALEXANDER V.: Knowledge-based support for complex systems exploration in distributed problem solving environments. In: *Communications in Computer and Information Science* Bd. 394 (2013), S. 147–161 — ISBN 9783642413599

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Knowledge Discovery in Big Data: Herausforderungen durch Big Data im Prozess der Wissensgewinnung am Beispiel des CRISP-DM

Thomas Göpfert[1], Andreas Breiter[2]

Abstract: Der Prozess valide, neuartige, potenziell nutzbare und verständliche Muster in Daten zu finden, wird als Knowledge Discovery in Database Prozess bezeichnet (KDD-Prozess). Die diesem Prozess zu Grunde liegende Datenbasis unterliegt einem ständigen Wandel. Doug Laney erkannte die Eigenschaften Volume, Variety und Velocity als neue Herausforderungen für IT-Organisationen. Heute werden diese Herausforderungen unter dem Begriff Big Data zusammengefasst. Die Auswirkungen von Big Data auf den KDD-Prozess sind bisher unzureichend untersucht. Ziel dieser Arbeit war es, die Herausforderungen durch Big Data im Prozess der Wissensgewinnung am Beispiel des CRISP-DM, eines der am meisten genutzten KDD-Prozessmodelle, zu analysieren. Durch ein systematisches Literaturreview wurden elementare Herausforderungen identifiziert und den Prozessschritten des Prozessmodells zugeordnet. Der überwiegende Teil der Ergebnisse konnten mittels Experteninterviews verifiziert werden. Neben der Identifikation zentraler Herausforderungen wurde deutlich, dass CRISP-DM bei der Analyse von Big Data Gültigkeit hat, aber zentrale Herausforderungen, vor allen in den Phasen der Datenvorverarbeitung, beachtet werden müssen.

Keywords: Big Data, Knowledge Discovery in Database, KDD, KDDM, CRSIP-DM, Data Mining

1 Einleitung

Aktuelle Studien gehen davon aus, dass die vorhanden Datenmenge (die Summe aller gespeicherten Daten) 2013 4,4 Zettabyte (1 Zettabyte sind 1021 Bytes) betrug. Es wird geschätzt, dass sich diese Datenmenge alle zwei Jahre verdoppelt. 2020 wird mit einem Datenvolumen von 44 Zettabyte gerechnet. IDC nennt diese Datenmenge in ihrer gleichnamigen Studie das „Digitale Universum“ und geht davon aus, dass diese Daten wie Dinge im physikalischen Universum auch, in ganz unterschiedlicher Form vorliegen [ID14].

Dieses Datenvolumen (Volume), die Datenvielfältigkeit (Variety) und die Geschwindigkeit in welcher Daten (Velocity) erzeugt werden, erkannte Doug Laney als neue Herausforderungen für IT-Organisationen [La01]. Heute bildet diese Beschreibung

[1] Universität Bremen, AG Informationsmanagement, Bibliotheksstraße 1 , 28359 Bremen, thg@informatik.uni-bremen.de
[2] Universität Bremen, Institut für Informationsmanagement Bremen GmbH, Am Fallturm 1, 28359 Bremen, abreiter@ifib.de

eine der geläufigsten Definitionen für Big Data [Ki14]. Die oben erwähnte Studie des IDC geht davon aus, dass 22 Prozent der Daten 2013 und 37 Prozent der Daten 2020 einen Nutzen bringen, wenn sie mit Schlagwörtern versehen und analysiert werden [ID14]. „Die intelligente Auswertung kann Organisationen wichtige Daten liefern. […] Es ist offensichtlich, dass solche Unternehmen einen Wettbewerbsvorteil erlangen, die aus der Vielzahl der Daten geschäftsrelevante Informationen filtern können“ [BI14]. Chen und Zhang beschreiben typische Anwendungen von Big Data in den Feldern Commerce and Business (Wirtschaft), Scientific Research (Wissenschaft) und Society Administration (Öffentliche Verwaltung) [PZ14]. Dabei macht nicht die Analyse der Daten an sich, sondern das Verstehen der Daten (d.h. u.a. das Anreichern der Daten mit Kontextinformationen), die Sicherstellung der Datenqualität und die Vorverarbeitung der Daten (zur Vorbereitung für die eigentliche Analyse) einen Großteil des Aufwandes bei der Datenanalyse aus. So nimmt Franks an, dass 95 bis 100 Prozent des Aufwands der gesamten Datenanalysen alleine darin besteht, diese Arbeitsschritte zu erledigen [Fr12].

Fayyad u.a. analysieren lange vor der intensiven Diskussion des Begriffes Big Data, dass Data Mining (also die eigentliche Anwendung von Algorithmen) nur ein Teil eines Prozesses ist, den sie Knowledge Discovery in Database (KDD-Prozess) nennen [FPS96]. Auch große Unternehmen haben dies erkannt und einen standardisierten Prozess zur Datenanalyse entwickelt [IB10]. Dieser Cross-Industry Standard Process for Data Mining (CRISP-DM) ist das heute am weitesten verbreitete Prozessmodell zur Datenanalyse [Be10].

Es ist anzunehmen, dass diese Prozessmodelle nicht die angesprochen Entwicklung der Herausforderungen von Volume, Variety und Velocity berücksichtigen. So entstanden diese Prozesse der Wissensgewinnung bereits in den Jahren 1996 (KDD-Prozess) und 2000 (CRISP-DM), während z.B. die meistverwendete Beschreibung der Eigenschaften von Big Data von Doug Laney auf das Jahr 2001 zurückgeht. Auf Grundlage dieser Annahme hat dieser Beitrag das Ziel, die Herausforderungen durch Big Data in den jeweiligen Phasen des Prozess der Wissensgewinnung am Beispiel des CRSIP-DM systematisch zu identifizieren und zu benennen.

2 Cross-Industry Standard Process for Data Mining

Der Cross-Industry Standard Process for Data Mining (CRISP-DM) ist ein industrie-, werkzeug- und anwendungsneutrales Prozessmodell. Es ist laut Berthold u.a. das am Weitesten verbreitete und eingesetzte Modell [Be10]. Der CRISP-DM 1.0 Step-by-step data mining guide beschreibt vier in ihrem Abstraktionslevel unterschiedliche Ebenen. Die obersten zwei Ebenen werden als CRISP-DM reference model bezeichnet, die zwei unteren Ebenen als CRISP-DM user guide. Jede Phase besteht aus generischen Aufgaben, welche zusammen die Ebene darunter bilden. Die Phasen mit ihren jeweiligen generischen Aufgaben bilden das CRIPS-DM reference model. Die Abstraktion dieser Phasen und Aufgaben ist so gewählt, dass diese für alle Data Mining Anwendungen

Gültigkeit haben (Vollständigkeit des Modells) und zukünftige Entwicklungen, z.B. neue Modellierungstechniken, diese Phasen und Aufgaben nicht beeinflussen (Stabilität des Modells) [IB10].

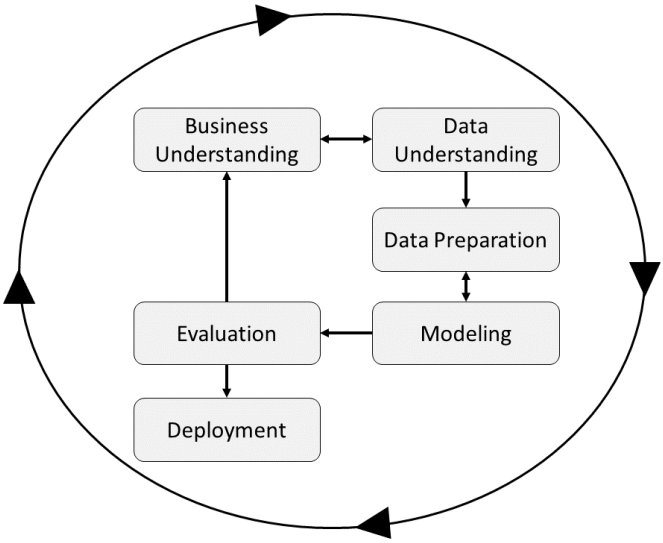


Abb. 1: CRISP-DM Modell (Eigene Darstellung nach [IB10])

Die Phasen des CRISP-DM reference model sind Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation und Deployment. Abbildung 1 gibt einen Überblick über den Zusammenhang dieser Phasen. Ziel der Phase Business Understanding ist es, ein Verständnis über das Projekt aufzubauen. Dazu zählen die Anforderungen aus Sicht des Kunden, die gegebenen Faktoren, welche zu berücksichtigen sind die Anforderungen für Data Mining Aufgaben, welche sich aus den Zielen des Kunden ableiten und ein erster Projektplan. Ein Verständnis der Daten zu erlangen, ist das globale Ziel der Phase Data Understanding. Hierfür werden die Daten gesammelt, beschrieben, erste Einblicke gewonnen, interessante Daten identifiziert, ggf. erste Hypothesen aufgestellt und die Qualität der Daten überprüft. Am Ende der Phase Data Preparation stehen die Datensätze, mit welchen in der Phase Modeling (sie sind Eingabe für die Data Mining Algorithmen) gearbeitet wird. In der Phase Modeling werden verschiedene Modellierungstechniken gewählt und angewandt. Die Ergebnisse der Modeling Phase werden mit den Zielen des Kunden in der der Phase Evaluation verglichen. Ausgehend von der Annahme, dass die Generierung und die Evaluation eines Modells nicht das Projekt abschließen, sondern vielmehr die Ergebnisse zum Zwecke der Wissensgewinnung aus Daten genutzt werden sollen, muss das Modell Anwendung finden, auch als Auslieferung (Deployment) bezeichnet. Hierzu müssen Ergebnisse ggf. in einer nutzbaren Form organisiert und dargestellt werden [IB10].

3 Stand der Forschung

Datensätze werden als Big Data bezeichnet, wenn es mit aktueller Technologie schwierig ist, diese zu erheben, zu pflegen, zu analysieren und zu visualisieren [PZ14], [Ja11]. Aktuelle empirische Studien des TDWI Germany e.V. „Big Data in deutschen Unternehmen – Relevanz und Reife“ [Gl12] oder der BITKOM „Big Data im Praxiseinsatz - Szenarien, Beispiele, Effekte“ [BI12] zeigen Herausforderungen bzw. Hürden bei der Einführung bzw. Umsetzung von Big Data Projekten in der Praxis. Dabei ist eine Mischung aus organisatorischen und technischen Faktoren festzustellen [Gl14].

Einige Beiträge untersuchen Big Data anhand oder mittels einzelner Prozessschritte, wie sie im KDD-Prozess oder im CRISP-DM vorkommen. So analysiert Rajpurohit, dass ein Hand-in-Hand zwischen Technologie und Business Understanding nötig und dazu ein Verständnis des KDD-Prozess Grundlage ist, um durch Big Data zu profitieren. Er führt dabei das Prozessmodell von Fayyad u.a. an und erklärt die einzelnen Schritte ohne dabei Modifikationen oder Herausforderungen bzgl. Big Data hervorzuheben. Allgemein stellt er allerdings deutlich die Schwierigkeit heraus, dass die Deckung der Kosten für die Einführung einer Big Data Lösung (Return of Investment) schwer zu kalkulieren sind [Ra13].

Auch Begoli und Horey stellen heraus, dass durch Big Data die Datenanalyse mittels KDD-Prozess die konsequente Lösung ist. Sie heben hervor, dass ein effektiver Einsatz dieser Methode allerdings technologische und organisatorische Erfahrung voraussetzt. Aus ihren praktischen Erfahrungen beim Oak Ridge National Laboratory (ORNL) heraus identifizieren sie drei wesentliche Design Prinzipien, deren Anwendung aus ihrer Sicht wichtig sind, um ökonomische, umfassende, flexible und sichere Lösungen für die Big Data Anforderungen im öffentlichen Sektor zu generieren. Diese Prinzipien sind: (1) die Unterstützung einer Vielzahl von Analysemethoden, (2) das Bereitstellen von Lösungen zur Speicherung und Verarbeitung von Daten in allen Prozessschritten und (3) die Verfügbarkeit und Verständlichkeit resultierender Daten [BH12].

Eine ausführliche Analyse von Big Data präsentieren Chen und Zhang in ihrer Arbeit „Data-intensive applications, challenges, techniques and technologies: A survey on Big Data“. Ein Teil dieser Arbeit besteht darin, die Herausforderungen durch Big Data anhand der Prozessschritte Datenerhebung und -speicherung, Datenübermittlung, Datenpflege, Datenanalyse und Datenvisualisierung zu beschreiben. Dabei sehen die Autoren in den Schritten Datenerhebung und Datenspeicherung und die Gewährleistung des Zugangs zu Big Data als eine Aufgabe mit top Priorität im Knowledge Discovery Prozess. Die Daten sollten einfach und sofort zugänglich sein. Die Vorverarbeitung (Datenbereinigung, Datentransformation, Datenkatalogisierung) der Daten vor der Analyse ist notwendig, um Data Mining Methoden anzuwenden. Auch für die eigentliche Datenanalyse müssen neue Methoden entwickelt werden. Nicht nur für die Speicherung, Pflege und Analyse, sondern auch für die Visualisierung werden laut Chen und Zhang neue Methoden benötigt [PZ14].

4 Methode

Basierend auf der Beobachtung, dass nach bestem Gewissen keine Analyse der Herausforderungen durch Big Data im CRISP-DM existiert, dass aber durchaus zahlreiche Herausforderungen durch Big Data benannt wurden, resultiert folgendes Vorgehen.

Zur Identifikation der bereits benannten Herausforderungen wurde ein systematisches Literaturreview, angelehnt an Kitchenham [Ki04], durchgeführt. Anhand der Forschungsfrage wurde eine Reviewstrategie abgeleitet und dokumentiert. Diese basiert auf den Phasen des CRISP-DM und wurde in sieben verschiedene Suchen (je Phasen plus eine allgemeine Suche nach Herausforderungen) mit unterschiedlichen Suchstrings aufgeteilt. Das Review wurde mittels der Suchmaschine „Web of Science“ der Firma Thomson Reuters[3] durchgeführt. Dabei wurden die Suchergebnisse auf das Erscheinungsjahr 2012-2014 und den zugeordneten Wissenschaftsbereich „Computer Science“ beschränkt. Die Suchergebnisse wurden nach Relevanz und Anzahl der Zitationen sortiert und eine entsprechende Anzahl an Treffern je Suche zur näheren Analyse ausgewählt. Es konnte 28 Arbeiten zur näheren Analyse herangezogen werden. Mittels zusammenfassender Inhaltsanalyse nach Mayring [Ma10] wurden Herausforderungen identifiziert, abstrahiert, generalisiert und den Phasen des CRISP-DM zugeordnet. Eine Zuordnung wurde vorgenommen, falls die Herausforderung eine Aufgabe der jeweiligen Phase tangiert. Darüber hinaus wurden übergreifende Herausforderungen gebündelt. Das daraus entstandene Kategoriensystem setzt sich dementsprechend aus deduktiven und einer induktiv gewonnenen Kategorie zusammen.

Aufbauend auf den Erkenntnissen aus dem systematischen Literaturreview wurden zwei leitfadengestützte Experteninterviews durchgeführt. Diese wurden transkribiert und ebenfalls anhand der zusammenfassenden Inhaltsanalyse und mittels des im systematischen Literaturreview entstandenen Kategoriensystems analysiert. Für die Interviews konnte ein Experte aus dem Bereich Wissenschaft und ein Experte aus dem Bereich Wirtschaft, also zwei der drei durch Chen und Zhang genannten Anwendungsbereiche von Big Data [PZ14] gewonnen werden. Im Bereich Wirtschaft konnte das Interview mit einem Projektleiter eines mittelständischen Unternehmens, spezialisiert auf die Beratung und Durchführung von Projekten im Bereich Business Intelligence, Data Warehouse und Berichtswesen durchgeführt werden. Im Bereich Wissenschaft wurde ein Sozialwissenschaftler mit dem Fokus auf die Analyse von Twitterdaten interviewt. Beide Interviewpartner führen oder führten bereits Big Data Projekte durch, die der genannten Definition von Big Data entsprechen.

[3] Webseite der Suchmaschine Web of Science: webofknowledge.com

5 Ergebnisse

Insgesamt wurden 208 elementare Herausforderungen mittels systematischem Literaturreview und Experteninterviews identifiziert und den Kategorien, also den Phase des CRSIP-DM reference model, zugeordnet. Im Folgenden werden die zentralen Herausforderungen in den einzelnen Phasen des CRISP-DM dargestellt. Eine jeweils einleitende Tabelle fasst wesentliche Herausforderungen zusammen und gibt an ob diese durch das systematische Literaturreview (im Folgenden Spalte L) oder durch die Experteninterviews (Spalte I) identifiziert wurden und/oder dem Stand der Forschung (Spalte F) entsprechen.

5.1 Business Understanding

Herausforderung	L	I	F
Business Case, Business Ziele und abgeleitete Anforderungen definieren	Ja	Ja	Ja
Zusammenarbeit Management und Kunden	Ja	Ja	Ja
Datenschutz und Ethik beachten	Ja, nicht Ethik	Ja	Ja, nicht Ethik
Projektplanung und Kostenschätzung/-planung	Ja	Ja, Kosten nein	Ja

Tab. 1: Zentrale Herausforderungen der Phase Business Understanding

Eine zentrale Aufgabe der Phase Business Understanding ist die Festlegung der Business Ziele (Aufgabe Determine Business Objectives). Eine erste Herausforderung dieser Phase stellt die Definition von Geschäftsszenarios dar, in welchen durch Big Data gewonnene nützliche Informationen eine Rolle spielen und einen Vorteil bringen. Ziele eines Big Data Projektes müssen klar definiert und insbesondere von allen Beteiligten verstanden werden. Diese Ziele sollten u.a. auch die Möglichkeiten aktueller Technologien berücksichtigen. Dazu muss auch der Kunde aktuelle Technologien verstanden haben. Big Data Projekte haben meist einen großen, ggf. experimentellen Charakter und bedürfen der Unterstützung des Managements. Es hat sich in den Interviews gezeigt, dass eng mit der IT-Abteilung des endgültigen Betreibers des Informationssystems zusammengearbeitet werden muss.

Die Ergebnisse des Literaturreviews und der Interviews ergeben auch ein einheitliches Bild in Fragen bzgl. des Datenschutzes. Der Schutz der Privatsphäre ist eine prozessübergreifende Herausforderung, betrifft aber insbesondere die Phase Business Understanding, in der entsprechende Regelungen und Fragestellungen mitgedacht werden müssen (Aufgabe Assess Situation). In der Phase Business Understanding müssen rechtliche Rahmenbedingungen erhoben und relevante Fragestellungen für die

nächsten Phasen vorbereitet werden. Der Interviewpartner im Bereich Scientific Research hob zusätzlich die Einhaltung ethischer Grundsätze als Herausforderung hervor.

Eine weitere Herausforderung ist die Planung des Projektes (Aufgabe Produce Project Plan). Datenverständnis zu erlagen und die Daten für die Analyse vorzubereiten benötigt einen Großteil der Zeit im Projekt. Im CRISP-DM entspricht dieses Datenverständnis der Phase Data Understanding, für welche im Projektplan entsprechend Zeit eingeplant werden sollte. Diese Einschätzung konnte sowohl bei der Literaturrecherche wie auch in den Interviews gewonnen werden. In den Interviews wurde hervorgehoben, dass auch das Verständnis der Technologie Zeit benötigt, sowohl bei den ExpertInnen, als auch beim Kunden. Die Schätzung und Planung des Aufwandes ist Voraussetzung für die Schätzung der Kosten. Diese stellen weite Herausforderungen dar, welche in der Literatur, in den Studien und in den Interviews gleichermaßen genannt werden.

5.2 Data Understanding, Data Preparation, Modeling

Herausforderung	**L**	**I**	**F**
Bestehende Speichertechnologien und Infrastrukturen sind zur Analyse nicht geeignet	Ja	Ja	Ja
Bestehende Analysesoftware zur Analyse nicht geeignet	Ja	Ja	Ja
Datenzugriff und Datenübertragung	Ja	Ja	Nein
Verstehen der Daten	Ja	Ja	Nein
Integration und Vorverarbeitung der Daten	Ja	Ja	Nein, nur Dauer Ladeprozess
Modellierungstechniken (Algorithmen) anpassen und nutzen	Ja	Ja	Ja

Tab. 2: Zentrale Herausforderungen der Phasen Data Understanding, Data Preparation und Modeling

Die Phasen Data Understanding, Data Preparation und Modeling beinhalten laut Franks einen Großteil des Aufwandes [Fr12] in einem Big Data Projekt und sind von übergreifenden technischen Herausforderungen beeinflusst. Zum einen sind die herkömmlichen ggf. bestehenden Technologien wie relationale DBMS, NAS Speichertechnologien oder herkömmliche Analyseverfahren wie OLAP oft nicht geeignet, um mit großen, unstrukturierten Datenmengen in vertretbarer Geschwindigkeit umgehen zu können. Zum anderen müssen neue, aktuelle Technologien durch ExpertInnen verstanden aber auch beherrscht werden. Die Erkenntnisse, dass aktuelle Technologien und das Verständnis dieser benötigt werden, konnte im systematischen Literaturreview gewonnen und durch die Experteninterviews belegt werden.

Die wesentlichen Herausforderungen beim Sammeln der Daten (Aufgabe Collect Initial Data) im Prozess der Wissensgewinnung sind der Datenzugriff und die

Datenübertragung. Der Datenzugriff muss verfügbar, erlaubt und sicher sein und ebenso den rechtlichen Anforderungen, wie dem Datenschutzrecht entsprechen. Datenübertragungsgegebenheiten müssen geprüft und dokumentiert werden. Durch eine Vielzahl von Datenquellen in einem Big Data Projekt kann das Management der Datenquellen und die Datenübertragung eine komplexe Herausforderung darstellen. Des Weiteren können Abhängigkeiten zu Schnittstellen (APIs) und darauf aufbauenden Werkzeugen vorhanden und der Zugriff mittels dieser ggf. eingeschränkt sein. Erfahrungen und Fachwissen bei der Benutzung von APIs und Werkzeugen wurden insbesondere im Interview im Bereich Scientific Research als Herausforderung hervorgehoben.

Die weiteren Aufgaben der Phase Data Understanding behandeln das Verstehen der Daten (Aufgaben Describe Data, Explore Data) und die Datenqualität (Aufgabe Vertify Data Quality). Dieses Verstehen (also das anreichern der Daten mit Kontextinformationen) ist mittels herkömmlicher Datenmodelle meist nicht möglich. Zum einen können Unsicherheiten in den Daten nicht abgebildet werden, zum anderen ist es nur schwer möglich unstrukturierte Daten zu modellieren. Aber nicht nur das Verstehen noch nicht modellierter Datenquellen, sondern auch das Verstehen bestehender Datenstrukturen ist eine Herausforderung, da vorhandene Datenmodelle meist unvollständig sind. Um ein Verständnis der Daten, unabhängig von den Modellen zu gewinnen, müssen Anfragen an diese gestellt werden. Aber auch hier stoßen herkömmliche Anfragetechniken, wie einfaches SQL oder zur herkömmlichen explorativen Datenanalyse eingesetzte Techniken wie OLAP, durch große, unstrukturierte Datenmengen an ihre Grenzen. Diese Art von Daten und die zusätzliche Tatsache, dass es sich meist um viele Datenquellen handelt, macht auch die Sicherstellung der Datenqualität zu einer Herausforderung. Für all diese Tätigkeiten müssen u.U. aktuelle Technologien, z.B. für die Erkundung großer verteilter, unstrukturierter Datenmengen oder die Überprüfung der Datenqualität von großen Datenmengen wie Apache Hadoop oder NoSQL Datenbanken eingesetzt werden.

Wurden die Daten gesammelt und verstanden, müssen diese vorverarbeitet und integriert werden (Phase Data Preparation). Diese Integration und Vorverarbeitung ist für die Erstellung der Modelle ein wichtiger Schritt. Daten müssen mit Informationen angereichert werden (z.B. Metadaten erzeugt oder mit Schlagworten versehen werden). Hierbei ist auch die Beachtung des Datenschutzes wichtig, denn durch gezielte Vorverarbeitungsschritte kann man diesem durchaus gerecht werden. Allerdings sind für all diese Tätigkeiten bei großen, unstrukturierten Datenmengen auch aktuelle Technologien nötig. Insbesondere bei der Benutzung verteilter Systeme ist eine intelligente Aufteilung der Daten notwendig, dies hebt der Interviewpartner im Bereich Commerce and Business mehrfach hervor. Die Vorverarbeitung und Integration der Daten wurde in diesem Interview als zentraler Unterschied zu herkömmlichen Datenanalyseprojekten identifiziert. Auch in der Literatur wurde die bedeutende Rolle dieser Phase hervorgehoben, z.B. [Fr12]. Auffällig ist, dass diese Herausforderungen, bis auf die erhöhte Dauer des Ladeprozesses, nicht in den genannten Studien als solche identifiziert wurden [Gl12], [BI12].

Bei der eigentlichen Analyse (Phase Modeling) ist es eine Herausforderung, die richtige Modellierungstechnik zu wählen. Da die Modellierungstechnik die Art der Vorverarbeitung der Daten bestimmt, sind diese Phasen durchaus mehrfach im Wechsel zu durchlaufen. Aus dem Stand der Forschung geht hervor, dass ein wichtiges Designziel in Big Data Projekten die Unterstützung einer Vielzahl von Analysemethoden ist [BH12]. NutzerInnen sollen die Möglichkeit haben Tools und Methoden frei zu wählen. Dies stellt im Prozess der Wissensgewinnung eine große, vor allem, technologische Herausforderung dar. Aber auch die Methoden, also auch die Algorithmen an sich, müssen ggf. weiterentwickelt und auf unstrukturierte, unsichere und verteilte Daten angepasst werden. Dabei muss auch Nebenläufigkeit, Skalierbarkeit und ggf. die Berechnung Vorort (bei den Datenquellen) unterstützt und aktuelle Technologien beachtet werden.

5.3 Evaluation und Deployment

Herausforderung	**L**	**I**	**F**
Evaluation von Modellen und Infrastruktur	Ja	Nein	Nein
Bestehende Speichertechnologien und Infrastrukturen sind zum Betrieb nicht geeignet	Ja	Ja	Ja
Bestehende Analysesoftware und Algorithmen sind im Betrieb nicht geeignet	Ja	Ja	Ja
Visualisierung von Big Data	Ja	Nein	Nein

Tab. 3: Zentrale Herausforderungen der Phasen Evaluation und Deployment

In der Phase Evaluation sollen die Modelle auf ihre Korrektheit überprüft werden. Dies ist durch die großen Datenmengen eine Herausforderung. Auch können Langzeitmodelle nur schwer unmittelbar verifiziert werden. Aber auch die eingesetzte Infrastruktur sollte evaluiert werden. Methoden hierfür sind kaum vorhanden. Auffällig ist, dass Herausforderungen betreffend der Evaluation nur in geringer Anzahl und ausschließlich mittels des Literaturreviews identifiziert werden konnten.

Ein anderes Bild ergab sich bei der Analyse der Herausforderungen in der Phase Deployment. Hier wird die Auslieferung des Modelles geplant. Da es sich in der Praxis, insbesondere durch das Interview im Bereich Commerce and Business unterstützt, meist um die Auslieferung des Analysemodelles an die IT- Abteilung des Kunden, d.h. die Überführung der Prozessergebnisse (Analysemodelle, Parameter der Modelle, Verfahrensweisen, Erfahrungen, usw.) in ein produktives Informationssystem handelt, sind hier viele Herausforderungen identifiziert wurden, die zum einen bei der Planung der Auslieferung auftreten, aber insbesondere bei der Auslieferung direkt. Die zentralen Herausforderungen lassen sich grob in die Bereiche: bestehende Speichertechnologie und Netzwerkinfrastruktur, bestehende Analysesoftware und Visualisierung untergliedern.

Die bestehenden Speichertechnologien und die Netzwerkinfrastruktur muss bei der

Planung der Auslieferung mitgedacht werden. Dabei sind Herausforderungen, dass die Anforderungen an Geschwindigkeit, Skalierbarkeit und Performance einer Big Data Analyse meist nicht von den bestehenden/herkömmlichen Speichertechnologien und Netzwerkinfrastrukturen unterstützt werden. Meist sind diese nicht für den verteilten Betrieb ausgelegt (bspw. hinsichtlich Fehlertoleranz, Konsistenz, Hochverfügbarkeit oder Nebenläufigkeit) und ebenso anfällig gegenüber inkonsistenten, unvollständigen oder ungenauen Daten. Beim Einsatz verteilter Systeme, z.B. ein Apache Hadoop Cluster, ist meist auch die vorhandene Netzwerkbandbreite ungenügend, wenn diese durch Map-Reduce-Jobs zusätzlich beansprucht wird.

Ein ähnliches Bild ergibt sich bei der Betrachtung herkömmlicher/bestehender Analysesoftware. Auch hier sind skalierbare Techniken, z.B. für Text Mining oder Soziale Netzwerkanalyse, nötig. Diese müssen ebenfalls parallel arbeiten und werden meist durch eine Middelware verwaltet. Eine Erweiterung bestehender Analysesoftware muss, unter Beachtung der Ergebnisse im Prozess, geplant werden.

Herausforderungen bei der Planung der Visualisierung konnten ausschließlich durch das systematische Literaturreview gefunden werden. Aber auch Begoli und Horey empfehlen, dass die resultierenden Daten verfügbar und verständlich sein müssen [BH12]. Herausforderungen bei der Visualisierung sind die Multidimensionalität der Daten und die Ungeeignetheit von herkömmlichen Tools und Techniken. Eine Quelle empfiehlt, dass Visualisierung völlig neu gedacht werden muss, anstatt sie an bestehende Ansätze anzupassen [PZ14].

6 Fazit

Durch ein systematisches Literaturreview und Experteninterviews wurden zahlreiche Herausforderungen durch Big Data identifiziert. Alle Herausforderungen konnten einer Phase des CRSIP-DM zugeordnet oder als übergreifende Herausforderung kategorisiert werden. Ein großer Teil der gefundenen Herausforderungen findet sich auch in empirischen Untersuchungen des TDWI Germany e.V. „Big Data in deutschen Unternehmen - Relevanz und Reife“ [Gl12] oder der BITKOM „Big Data im Praxiseinsatz - Szenarien, Beispiele, Effekte“ [BI12] wieder. Auffällig ist, dass Herausforderungen in den Phasen Data Understanding und Data Preparation, also den Phasen die laut Franks einen Großteil des Aufwandes ausmachen [Fr12] in den genannten empirischen Untersuchungen nicht als Herausforderungen von Unternehmen genannt werden. Dies gilt ebenfalls für Herausforderungen der Evaluation und Visualisierung. Dagegen ist die Mischung aus organisatorischen und technischen Herausforderungen sowohl Ergebnis dieser Untersuchung, als auch der genannten Studien.

Durch die Beschreibung der Herausforderungen in den einzelnen Phasen des CRISP-DM wurde deutlich, dass die Planung und der Einsatz aktueller Technologien entscheidend und eine der wesentlichen Herausforderungen im Prozess sind. Dies findet sich ebenso

in den Ergebnissen des systematischen Literaturreviews als auch in den Experteninterviews mehrfach wieder. Dies ist nicht verwunderlich, da Datensätze eben dann als Big Data bezeichnet werden, wenn es mit aktuellen Technologien schwierig ist, diese zu erheben, zu pflegen, zu analysieren und zu visualisieren [PZ14], [Ja11]. Begoli und Horey erkannten das Bereitstellen von Lösungen zur Speicherung und Verarbeitung von Daten in allen Prozessschritten als ein zentrales Designziel in Big Data Projekten [BH12]. Ein Interviewpartner hob hervor, dass eine extra Phase zum Technologieverständnis im Projekt durchgeführt werden musste und diese das Projekt verzögerte. Dies lässt den Schluss zu, dass Technologieentscheidungen frühzeitig, d.h. in der Phase Business Understanding, mitgedacht werden müssen, um die Grundlage späterer Phasen zu legen und diese nicht zu verzögern. Diese ist bisher so im Prozessmodell CRISP-DM nicht im Detail vorgesehen, kann aber durchaus in Aufgaben wie Assess Situation oder Produce Project Plan hineininterpretiert werden. Weiterhin konnte gezeigt werden, dass viele Herausforderungen bei der Auslieferung an sich auftreten. Hierfür ist in CRISP-DM nur die Planung vorgesehen und es bleibt fraglich, ob dies die Problematik ausreichend abdeckt. Es wurde deutlich, dass der Einsatz eines Prozessmodells zur Wissensgewinnung aus Big Data möglich und sinnvoll ist. Allerdings sollten die aufgezeigten Herausforderungen der einzelnen Phasen beachtet und für das individuelle Projekt eingeplant werden. Dabei sollte, neben der Planung und dem durchgehenden Einsatz von aktuellen Technologien, ein besonderer Fokus auf den Phasen Data Understanding und Data Preparation liegen.

Literaturverzeichnis

[Be10] Berthold, M. R. et al.: Guide to intelligent data analysis. How to intelligently make sense of real data. Springer, London, 2010.

[BH12] Begoli, E.; Horey, J.: Design Principles for Effective Knowledge Discovery from Big Data: 2012 Joint Working IEEE/IFIP Conference on Software Architecture (WICSA) & European Conference on Software Architecture (ECSA), 2012; S. 215–218.

[BI12] BITKOM: Big Data im Praxiseinsatz - Szenarien, Beispiele, Effekte. http://www.bitkom.org/files/documents/BITKOM_LF_big_data_2012_online%281%29.pdf, 09.02.2015.

[BI14] BITKOM: Leitfaden Big-Data-Technologien - Wissen für Entscheider. Leitfaden. http://www.bitkom.org/files/documents/BITKOM_Leitfaden_Big-Data-Technologien-Wissen_fuer_Entscheider_Febr_2014.pdf, 18.04.2015.

[FPS96] Fayyad, U. M.; Piatetsky-Shapiro, G.; Smyth, P.: From Data Mining to Knowledge Discovery: An Overview. In (Fayyad, U. M. et al. Hrsg.): Advances in knowledge discovery and data mining. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1996; S. 1–34.

[Fr12] Franks, B.: Taming the big data tidal wave. Finding opportunities in huge data streams with advanced analytics. John Wiley & Sons Inc, Hoboken, New Jersey, 2012.

[Gl12] Gluchowski, P.: Big Data in deutschen Unternehmen - Relevanz und Reife.

www.tdwi.eu/wissen/studien, 01.05.2015.

[Gl14] Gluchowski, P.: Empirische Ergebnisse zu Big Data. In HMD Praxis der Wirtschaftsinformatik, 2014, 51; S. 401–411.

[IB10] IBM: CRISP-DM 1.0 Step-by-step data mining guide. http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=YTW03084USEN, 10.12.2014.

[ID14] IDC: Executive Summary: Data Growth, Business Opportunities, and the IT Imperatives. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm, 18.04.2015.

[Ja11] James Manyika; Michael Chui; Brad Brown; Jacques Bughin; Richard Dobbs; Charles Roxburgh; Angela Hung Byers: Big data: The next frontier for innovation, competition, and productivity. http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx, 18.04.2015.

[Ki04] Kitchenham, B.: Procedures for performing systematic reviews. Keele University, Keele, 2004.

[Ki14] Kitchin, R.: The Data Revolution. Big Data, Open Data, Data Infrastructures and Their Consequences. SAGE Publications, London, 2014.

[La01] Laney, D.: 3D Data Management Controlling Data Volume, Velocity, and Variety. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf, 13.10.2014.

[Ma10] Mayring, P.: Qualitative Inhaltsanalyse. Grundlagen und Techniken. Beltz, Weinheim, 2010.

[PZ14] Philip Chen, C. L.; Zhang, C.-Y.: Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. In Information Sciences, 2014, 275; S. 314–347.

[Ra13] Rajpurohit, A.: Big data for business managers — Bridging the gap between potential and value: 2013 IEEE International Conference on Big Data, 2013; S. 29–31.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

How to make quantitative data on the web searchable and interoperable part of the common vocabulary

Wolfgang Orthuber[1]

Abstract: The elements of the common vocabulary (e.g. words) are quickly known by all participants of a conversation. It was possible to extend on the web the vocabulary effectively by HTTP URLs, because their content is quickly viewable. A further extension is possible, which is powerful but not used up to now: Elements (vectors) of metric spaces, identified by a HTTP URL *plus* a sequence of values, where the URL locates the definition of the metric space. All elements and with this all well defined quantitative data can become searchable and quickly viewable and so part of the common vocabulary. In this paper we introduce the concept and propose worldwide multidimensional "Domain Spaces", which can be defined online by all users and which can be efficiently used as container of a huge extended vocabulary, especially user defined searchable quantitative data of all kinds.

Keywords: Metric Space, Domain Space, DS, Domain Vector, DV, Searchable quantitative data, Vocabulary, Standards, Search process, Selection process, Distributed data structures, Feature evaluation and selection, Feature Vector, Metric Search, High Resolution Search, Numeric Search, Interoperable quantitative data

1 Introduction and state of the art

The development of the here shown approach for viewable, searchable and interoperable quantitative data on the web has been started already in 2006 and meanwhile is demonstrated online [Or6]. Though the concept is designed for efficient handling of quantitative data, it is not restricted to these. [Or5] contains a detailed description and this paper introduces the concept. Easy definitions of multidimensional searchable data by users and adapted search interfaces are considered as important. Identification makes quantitative data interoperable, searchable and efficiently evaluable. With this they become also quickly viewable and so a substantial extension of the common searchable vocabulary. We hope this also leads to a constructive discussion about efficient structured and identified multidimensional data on the web.

1.1 Problem

The term "common vocabulary" is not uniquely defined. Therefore we define it initially in a reasonable way. Most important is that every element of the common vocabulary is

[1] University Medical Center Schleswig-Holstein, Department of Orthodontics, Arnold Heller Str. 3 House 26, 24105 Kiel Germany, orthuber@kfo-zmk.uni-kiel.de

quickly known by conversing people. We know that clickable information was very successful and became even defining characteristic of the web, because it can be opened (usually within 1 second) and so viewed quickly. Therefore by "common vocabulary" we mean all *quickly recognizable information* on the web, especially well known words, symbols and clickable information. At this information is "clickable", if in case of appropriate hardware and software after clicking on a text or image a window with more detailed descriptive information opens within 1 second.

Well defined resources often have well defined quantitative properties or features. These features fully depend on the resource, and usually describe the resource more precisely. For example we know that a "car" has important quantitative features which describe it precisely. Also conventional words of language can have specific quantitative features, e.g. "color" can have three percentual parts of (by wavelength well defined) "red", "green" and "blue". Up to now there is no systematic concept which allows to combine text and other clickable resources on the web with individual multidimensional quantitative data in a way that after click the quantitative data are shown by definition or even graphically, so that they are quickly recognizable. The kind of data should be also recognizable by software (machine readable), so that they become searchable.

1.2 Current approaches, shortcomings

First we look at the current state of the art. There are approaches for representation of data on the web [McGui][W3RDF][HL7F] including quantitative data, but the found approaches do not allow slender and efficient representation of multidimensional data and no definition of world wide multidimensional (metric) spaces. This can be shown best using examples from latest literature, e.g. in medicine where quantitative data are important and decision relevant. The FHIR specification [HL7F] is a next generation standards framework created by HL7 [HL7]. At this interoperable data are coded in resources. We look at an exemplary resource example:

```
<Observation xmlns="http://hl7.org/fhir">
  <text>
    <status value="generated"/>
    <div xmlns="http://www.w3.org/1999/xhtml">
     Jan 30 2014: Body Weight = 185 lbs</div>
  </text>
  <name>
    <coding>
      <system value="http://loinc.org"/>
      <code value="3141-9"/>
      <display value="Weight Measured"/>
    </coding>
  </name>
  <valueQuantity>
    <value value="185"/>
    <units value="lbs"/>
    <system value="http://unitsofmeasure.org"/>
    <code value="[lb_av]"/>
  </valueQuantity>
```

```
</Observation>
```

Fig. 1: Excerpt of http://www.hl7.org/fhir/observation-examples.html [HL7FEX] which codes the observation of body weight.

"http://hl7.org/fhir" in Fig. 1 specifies the namespace. According to [W3N] "It is not a goal that it be directly usable for retrieval of a schema (if any exists)." The URL "http://hl7.org/fhir" contains a human readable description of the general FHIR concept, *somewhere* in it the resource "observation" is defined. The URL "http://loinc.org" combined with the code "3141-9" yields the URL http://loinc.org/3141-9 where the quantity "Body weight Measured" is defined in human readable form. The unit of measurement is not determined there, so it is specified subsequently together with the value "185" as "lbs" with http://unitsofmeasure.org and "[lb_av]". Fact is: The more defining information is necessary in the data, the less standardized are the data. So (varying) defining information together with data is unnecessary and even disadvantageous.

1.3 Proposed approach

Fig. 1 contains defining information (e.g. about the unit) but no URL to a standardized definition which is machine readable in uniform format. In current interoperability approaches we have not found usage of such *online definitions*. This means, that together with the (*variable*) data a URL is given which points to a complete (*constant and reliable*) definition in standardized machine readable form. The data are given as sequence of values (variables). They are following after the URL of their definition, there is no further defining or redundant information. The online definition must be reliable and complete, it can contain also extended information in multiple languages. We can use for the online definition a standardized form, e.g. HTML [MF] or in XML. The subsequently described DS definitions (see 3.1) are the online definitions which we propose. Fig. 5 shows an abbreviated example of a definition and Fig. 2 exemplary data. It contains the URL of such a online (DS) definition plus 2 numbers and represents the values of Fig. 1 in machine readable form.

```
<v http://numericsearch.com/bw.xml; 2014-01-30; 83.914>clickable text</v>
```

Fig. 2: Values of Fig. 1 in exemplary representation using the URL of an online definition which contains the definition of a metric space (DS, see 2.1). The complete line can be seen as element (DV, see 2.2) of this space. The online definition (DS definition) can represent very much information in standardized way, e.g. default clickable texts in different languages.

Fig. 2 shows on a single line a possible example of a short new standard (DV, see 2.2) which could represent the date and body weight of Fig. 1. Here 83.914 represents the weight (185 lbs) in kg (we recommend usage of SI units in basal dimension definitions). The compactness of Fig. 2 in comparison to Fig. 1 makes it practicable to place multidimensional quantitative data into HTML pages, and because the URL points to an online definition which completes the information of the quantitative data, after click the complete information can be displayed in human readable form. The approach can be

generalized for all multidimensional data. We will deepen details subsequently.

2 Data structures

Fig. 2 shows a HTTP URL plus a sequence of values. To make it searchable we can understand every value as a dimension of a metric space. Every pair of elements of a metric space has a well defined nonnegative distance which is the smaller, the greater the similarity is. Therefore in our approach multidimensional quantitative data (Fig. 2) are defined as elements (vectors) of one- or multidimensional metric spaces. This allows similarity search by minimizing the distance. To cover all interesting topics, all users must get the practicable possibility to compose own online definitions of metric spaces. Subsequently we call these metric spaces "Domain Spaces" (DSs) and their elements "Domain Vectors" (DVs). We selected these names because DSs can be defined by web users according to their domain of interest, and because of their applicability in Domain Ontologies (see [Haas]).

2.1 Domain Space (DS)

Fig. 3 shows the structure of a DS definition. Main content is an ordered sequence of dimension definitions. The DS definition and every dimension definition can be addressed using a HTTP URL.

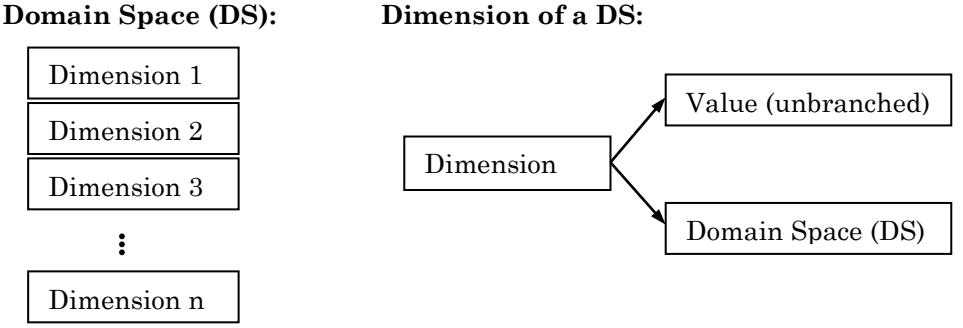


Fig. 3: Structure of a DS and DS dimension. The DS and every of its dimensions have a unique name (HTTP URL). Every dimension of a DS can represent an unbranched value or again a DS.

Fig. 3 shows that DS dimensions can represent a (unbranched) value or again a DS. The value is numeric by default. As special feature a value can also represent text or binary data using e.g. the discrete metric as distance function. If the dimension represents a DS, it is called branched. Because a dimension (of a DS) again can represent a DS, DS definitions can be nested and reused within new DS definitions.

2.2 Domain Vector (DV)

For identification of DVs (quantitative data) on the web we need a unique identifier (URI). It is efficient to use for this a Uniform Resource Locator (URL) [W3U] of the definition of the containing DS. This URL identifies the data *and* simultaneously locates their definition. Therefore the DV content starts with the HTTP URL of the DS definition. After this identifier the list of values follows usually in the same order like in the DS definition. So we get the above mentioned compact structure *HTTP URL plus a sequence of values* of a DV (data) on the web, as shown in Fig. 2. Additional identification of values by (abbreviated) HTTP URLs of dimension definitions is possible. Identification is necessary if the order is changed, e.g. if dimensions are bypassed.

3 Implementation

We made an online implementation which demonstrates on a local database important key steps: Guided definition of DSs and DVs, interoperable data formats which unambiguously depend on the definition, text search within parts of definitions, e.g. for selection of DSs for numeric search. Most interesting is multidimensional similarity search with and without additional conditions, where the search window is automatically adapted to the chosen DS (e.g. Fig. 7 and Fig. 9). In this paper we will show important steps and results of the procedure.

3.1 Definition of a DS

It is necessary that users can define own DSs. For demonstration purposes we show in Fig. 4 the definition of the unbranched simple space "BodyWeight". Details of every dimension can be defined online.

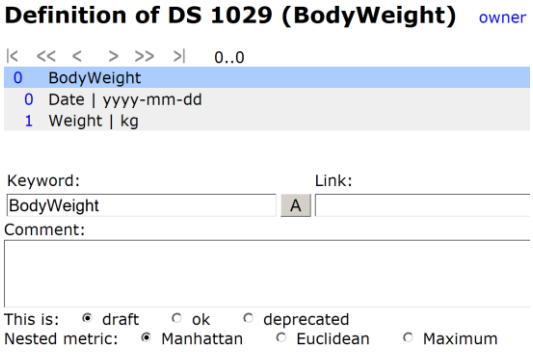


Fig. 4: Definition of a simple unbranched DS "BodyWeight".

Fig. 5 shows a possible XML representation of this DS definition [OrExDS].

```
<DS>
  <kw>BodyWeight</kw>
  <dim>
    <format>yyyy-mm-dd</format>
    <kw>Date</kw>
    <kw>yyyy-mm-dd</kw>
  </dim>
  <dim>
    <format>float</format>
    <kw>Weight</kw>
    <kw>kg</kw>
  </dim>
</DS>
```

Fig. 5: Possible abbreviated short XML representation of the definition of the DS "BodyWeight". This is only an example for explanation purposes. A later standard will deviate from this.

If "http://numericsearch.com/bw.xml" is the URL of this DS definition (Fig. 5), then Fig. 2 could represent a possible DV of this DS. It could be displayed on the screen as "clickable text" similarly like a clickable hyperlink. After click an adapted browser can show a detailed representation of the quantitative data, also our online implementation demonstrates this (Fig. 6) [OrExDV].

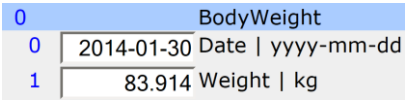


Fig. 6: Exemplary representation of the DV (quantitative data) of Fig. 2 [OrExDV] using the DS definition (Fig. 4).

Fig. 7 shows an exemplary search (search window online [OrExDV]) in a small DS "BodyWeight" with 4 DVs.

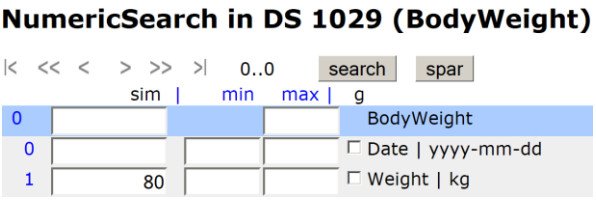


Fig. 7: Possible search within DS "BodyWeight" [OrExS1], Fig. 8 shows the search result.

```
    d   a
    0| 15 o clickable-text-1 | 2014-02-05, 80,
3.914| 24 o clickable-text | 2014-01-30, 83.914,
    5| 21 o clickable-text-2 | 2014-03-10, 75,
   10| 15 o clickable-text-3 | 2014-04-20, 70,
```

Fig. 8: Search result of Fig. 7. Column "a" shows the access count to the DV, Column "d" the distance (Manhattan metric with weight 1), and the last column the last dimension of the DV (weight).

Fig. 8 demonstrates the principle of similarity search in this simple example. As distance function default Manhattan Metric (in case of one dimension the absolute difference) is chosen. Let w represent the last dimension (weight) in a DV, d the distance. Because weight 80kg was searched, d=|80-w|. The search result is sorted that the rank is the higher, the smaller d is.

4 Evaluation

A DV represents quantitative data by a HTTP URL plus a sequence of values (Fig. 2). It is interoperable because data of the same kind have the same format and can be recognized by a unique URL. It is clickable like today a HTTP URL is clickable, if the URL points to a complete standardized definition of the values (a DS definition, e.g. Fig. 5) and the browser is programmed to combine it with the values. Also our implementation demonstrates clickable text and after click combination of values with a DS definition (Fig. 6).

Less trivial is search of DVs, especially its practicability. Because on the web all DVs of the same DS have the same identifier (the URL of the DS definition), a crawler can collect these into the same database. So DVs are searchable. But it is not clear to what extent search of DVs is possible in larger DSs. Therefore we generated a high dimensional nested DS and filled it with 100001 DVs. It combines according to 2.1 unbranched DSs. Two of them (a 4-dimensional DS with Euclidean Metric and a 140-dimensional DS with Manhattan Metric) were filled with pseudo random numbers which are equally distributed between 0 and 10. To demonstrate the effect of similarity search we searched in a 2 dimensional subspace (Fig. 9).

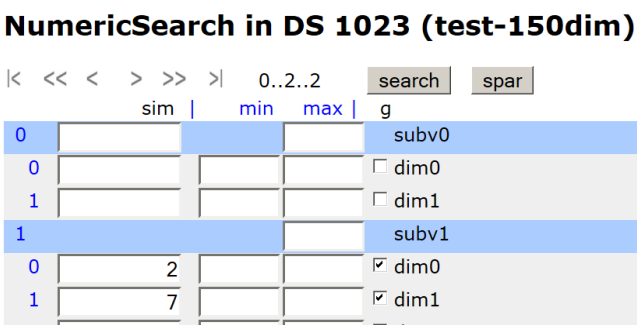


Fig. 9: Excerpt of the search window, Fig. 10 shows the search result.

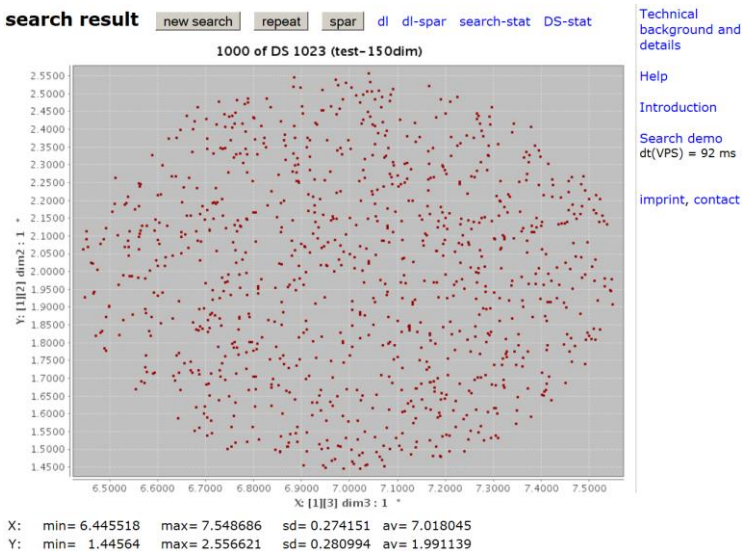


Fig. 10: Search result of Fig. 9, shown are the 1000 found DVs of 100001 DVs.

Fig. 10 shows the graphical result after search of Fig. 9. The content of the searched dimensions of the 1000 nearest of 100001 DVs is displayed. Due to the Euclidean Metric an elliptic shape around the searched coordinates results, because it represents in the given scale geometrically the set of nearest points around the point with the searched coordinates.

Now we test the performance of multidimensional similarity search in a more complex example with combined Euclidean and Manhattan metric. Let x_m denote a coordinate of a DV within the Euclidean Space and xs_m the corresponding searched coordinate and y_n a coordinate of a DV within the Manhattan Space and ys_n the corresponding searched coordinate, then the distance d is: $d = \sqrt{\sum_{m=1}^{4}(x_m - xs_m)^2} + \sum_{n=1}^{n\max}|y_n - ys_n|$. (1)

Using this distance function we performed similarity search with $n\max \in \{0,10,20,30,40,50,60,70,80,90,100,110,120,130,140\}$. We want to emphasize that due to the curse of dimensionality [Ag] high dimensional similarity search (nmax>10) tends to become inefficient in case of independent dimensions, else (in case of dependent dimensions) dimension reduction techniques are recommendable (e.g. [Indyk]). If the dependencies of dimensions are hidden, initially this can be difficult. High dimensional search was tested here to examine the performance of the system also in case of difficult situations.

According to (1) the total count of searched dimensions is nmax+4. Fig. 11 shows the average search times in case of 4..144 searched dimensions. The implementation was programmed using java jdk-7u4-linux-x64.rpm, apache-tomcat-7.0.27.zip, synchronized

index [Or5]. To save costs this implementation was done on a virtual server. It uses 14GB RAM and 4 cores of an Intel Core i7 CPU with 3.2 GHz, 1000 GB HDD with Linux Cent OS 6.3.

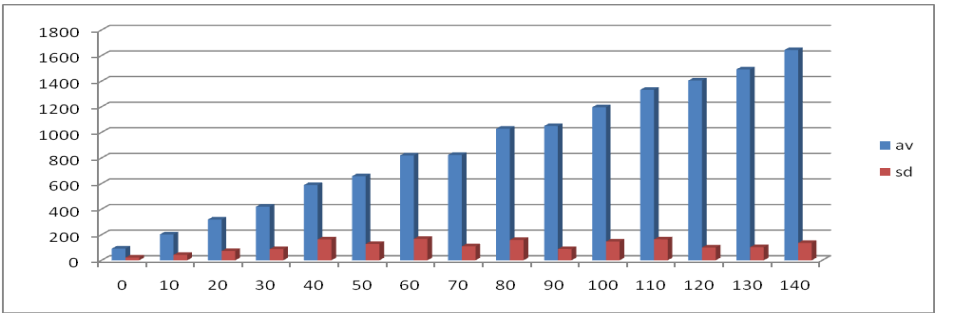


Fig. 11: Horizontal axis: nmax (=Count of searched dimensions in Manhattan Space)
Vertical axis: av (=average search time of 20 similarity searches within 100001 DVs in milliseconds) and sd (=standard deviation of search time)

Fig. 11 shows that it is possible to search through 100001 DVs on the average within 100ms in case of 4 dimensions, within 200ms in case of 14 dimensions and within 1700ms in case of 144 dimensions. Independently of this search performance could be enhanced by parallel processing, which is possible after splitting the space and reunion of the search results.

5 Discussion

According to [W3LT] at the June 2011 Semtech conference it was announced that a query of 1,009,690,381,946 triples was done in 338 hours for an average rate of 829556 triples per second. The driving force has been Amdocs and their AIDA platform. So averagely per 100ms 82956 triples have been processed. In another source [Erl] was stated "Loading speed for data in the Turtle syntax is up to 38000 triples per second." (i.e. 3800 triples within 100ms) The here presented implementation searched in case of 4 dimensions on the average within 100ms through 100001 DVs, using a cheap virtual server. So it becomes clear that DV search is practicable and very efficient compared to existing approaches.

The in Fig. 3 shown structure of DSs definitions is similar to that of an ordered directory tree: A DS combines dimensions which represent values or again DSs (like a directory contains files or again directories). At this in a DS definition the order of dimension definitions is important because it determines the order in which values can be given in a DV without necessary dimension identifier. Moreover in case of DSs circular definitions are possible, i.e. a DS can contain somewhere in its own tree a dimension which represents (points back to) an own definition. Therefore software has to stop after a predefined count of circular extensions.

"Why does this extend the searchable vocabulary?" is a natural question due to the topic and explicitly answered: The extension of the common searchable vocabulary by quantitative data means, that these are quickly viewable (quickly known) and searchable by readers. Fig. 6 shows an example for this. With this not only the clickable text (e.g. "BodyWeight") but also underlying quantitative data become part of the quickly viewable and searchable common vocabulary. The definitions of the numbers and names can be translated also into other languages, so that after clicking on a DV the definition can be shown in the language of the reader (if known by the system). Words within conventional text could be used as clickable text of a DV (Fig. 2) and get additional searchable quantitative features, e.g. "drive" could get the additional feature "speed in km/h". If wished, such "quantified text" could contribute to more precise conversation without being more difficult to read than conventional text. We hope the scope of such an extension of conventional language (for technical, medical, legislative texts etc.) will be recognized. A detailed discussion [Or5] would exceed the scope of this paper, here some notes:

- Syntax for definition of DSs and DVs must be standardized.
- Definition of a DS shows data which are interesting in a certain domain. This increases motivation for writers to provide these data, to make the web more informative.
- The Definition of a DS (with quantitative dimensions) defines all elements with their connection and with this much more than the definition of a word. Therefore range and resolution of DV based description and search is much higher than word based description and search.
- Though the definition of a DS can be very elaborate, patents on DS definitions are counterproductive. DS definitions and DVs can be seen as part of language and of course patents on parts of language should not be possible.
- It is important that DS definitions are stable and reliable. The owner of a certain domain can provide some warranty, moreover DS definitions can be stored by search engines to detect possible incompatible changes.
- Redundant definitions should be avoided or linked together with equivalent definitions using "sameAs" directives [W3Ont]. For this focused text search within DS definitions is necessary.
- It possible to derive from every DS automatically an "evaluation DS" in an open database to give users the possibility to evaluate every DV of the DS with an own "evaluation DV" of the "evaluation DS".
- DSs with textual dimensions allow construction of sentences with given structure, also triples of the semantic web [BeSe].
- DSs whose dimensions describe preconditions (1), a following decision (2) and the following results (3) can be used for decision support e.g. in medicine.

6 Conclusion

Introduction of online defined standardized metric spaces (DSs resp. "Domain Spaces") is an efficient means for extending the common vocabulary by user defined quantitative data (DVs) and for making quantitative data interoperable and searchable. Due to the far reaching technical potential it is recommendable to deepen this topic by science and economy.

Literature

[Ag] Aggarwal, C. Hinneburg, and A. Keim, D.: On the Surprising Behavior of Distance Metrics in High Dimensional Space. ICDT 2001, Vol. 1973, 420-434. DOI: http://dx.doi.org/10.1007/3-540-44503-x_27, 2001.

[An] Andoni, A. and Indyk, P.: Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. Communications of the ACM, vol. 51, no. 1, 2008, p. 117-122. http://mags.acm.org/communications/200801/#pg119, 2006.

[BeSe] Berners-Lee, T., .Hendler, J., and Lassila, O. 2001. "The semantic web." Scientific american 284.5, 28-37.http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html

[Erl] Erling, O., & Mikhailov, I. RDF Support in the Virtuoso DBMS. In Networked Knowledge-Networked Media (pp. 7-24). Springer Berlin Heidelberg. 2009.

[Gae] Gärdenfors, P. Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge, MA, 2000.

[Haas] Haas, P.: Medizinische Informationssysteme und Elektronische Krankenakten. Springer, Berlin, Heidelberg, 2005.

[HL7] HL7 Health Level Seven® INTERNATIONAL, http://www.hl7.org/, viewed June 2015

[HL7F] HL7 INTERNATIONAL: FHIR, http://www.hl7.org/fhir/ , viewed June 2015

[HL7FEX] HL7 INTERNATIONAL: Resource Observation - Examples, Simple Weight Example, http://www.hl7.org/fhir/observation-examples.html , viewed June 2015

[Indyk] Indyk P.: Dimensionality Reduction Techniques for Proximity Problems, Stanford University. (650-723-4532). http://people.csail.mit.edu/indyk/proxy99.ps, 1999.

[McGui] McGuinness, D. L., & Van Harmelen, F. OWL web ontology language overview. W3C recommendation, 10(10), 2004.

[MF] Microformats: Get started, http://microformats.org/wiki/get-started, viewed June 2015

[Na] Nakao M. A., Axelrod S.: Numbers are better than words. Verbal specifications of frequency have no place in medicine. Am J Med 1983, 74, 1061–1065, 1983.

[Or1] Orthuber, W., Fiedler, G., Kattan, M., Sommer, T., Fischer-Brandies, H.: Design of a global medical database which is searchable by human diagnostic patterns. The Open Medical Informatics Journal 2, 21-32, 2008.

[Or2] Orthuber, W., & Dietze, S.: Towards Standardized Vectorial Resource Descriptors on the Web. In GI Jahrestagung (2) (pp. 453-458), 2010.

[Or3] Orthuber, W., Papavramidis E.: Standardized Vectorial Representation of Medical Data in Patient Records, Medical and Care Compunetics 6, 153-166, 2010.

[Or4] Orthuber, W.: General approach to similarity search of resources with numeric features on the web. Open Data on the Web. http://www.w3.org/2013/04/odw/papers#al1, 2013.

[Or5] Orthuber, W.: Worldwide Domain Spaces make quantitative data searchable and prepare these for interoperable exchange, http://numericsearch.com/wwdomainspaces.pdf, viewed 15.06.2015

[Or6] Orthuber, W.: NumericSearch - online implementation of vectorial description and search, http://NumericSearch.com, viewed 24.04.2015

[OrExDS] Orthuber, W.: Temporary DS example, http://numericsearch.com/wcomedit.jsp?i9=77, viewed 14.06.2015

[OrExDV] Orthuber, W.: Temporary DV example, http://numericsearch.com/wcomedit.jsp?i9=72&i7=1029&i4=0, viewed 14.06.2015

[OrExS1] Orthuber, W.: Temporary search example 1, http://numericsearch.com/w7s.jsp?i7=1029, viewed 14.06.2015

[OrExS2] Orthuber, W.: Temporary search example 2, http://numericsearch.com/w7s.jsp?i7=1023, viewed 14.06.2015

[W3M] W3C: Metric Spaces on the Web, Community Group, in https://www.w3.org/community/numericweb/, viewed May 2015.

[W3N] W3C: Namespaces in XML 1.0 (Third Edition), W3C Recommendation 8 December 2009, in http://www.w3.org/TR/REC-xml-names/ , 2009

[W3Ont] W3C: OWL Web Ontology Language, W3C Recommendation, owl:sameAs, in http://www.w3.org/TR/owl-ref/#sameAs-def, 10 February 2004

[W3LT] W3C: LargeTripleStores, AllegroGraph, http://www.w3.org/wiki/LargeTripleStores , viewed June 2015

[W3U] World Wide Web Consortium: URIs, URLs, and URNs: Clarifications and Recommendations 1.0, http://www.w3.org/TR/uri-clarification/, 2001

[W3RDF] World Wide Web Consortium. RDF 1.1 Concepts and Abstract Syntax. http://www.w3.org/TR/rdf11-concepts/, 2014

[Za1] Zander S. and Schandl B. Context-driven rdf data replication on mobile devices. Semantic Web Journal Special Issue on Real-time and Ubiquitous Social Semantics, 1(1), 2011

[Zezula] Zezula, P., Amato, G., Dohnal, V., Batko, M. Similarity Search. The Metric Space Approach. Series: Advances in Database Systems, Vol. 32., Springer, Berlin, Heidelberg. 2005.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Cognitive and Dependable Systems

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Workshop for Cognitive and Dependable Systems

Douglas Cunningham[1], Ronald Römer[2]

Abstract: Im Zuge der digitalen Revolution sind Computer und andere Geräte der Informations- und Kommunikationstechnologie aktiv an nahezu jeder Facette des modernen Lebens beteiligt - von der automatischen Steuerung von Haustechnik, Automobilen und Flugzeugen über eine semiautonome Kontrolle von Kraftwerken, Energie- und Verkehrsnetzen bis hin zur Medizin- und Rehabilitationstechnik und zu Assistenzsystemen, um nur einige Beispiele zu nennen. Computer und Algorithmen werden nicht nur immer leistungsfähiger, ressourcenschonender und im Allgemeinen smarter, sondern auch immer zugänglicher für den Menschen und ermöglichen so ein effizienteres Arbeiten. Anders ausgedrückt: Computer führen menschliche Tätigkeiten so aus, wie wir es tun würden und sie haben begonnen, auf menschliche Art und Weise mit uns zu kommunizieren.

Der Trend hin zu dieser ”smarten” Technologie ist das Ergebnis zweier parallel verlaufender Entwicklungen. Auf der einen Seite haben kognitive und experimentelle Psychologen, Semiotiker und Logiker Modelle zur Beschreibung menschlicher Tätigkeiten einschließlich der natürlichen Kommunikation etabliert. Obwohl diese Modelle oft auf einer großen Menge von Daten basieren, sind sie normalerweise lediglich deskriptiv. Auf der anderen Seite benutzen Informatiker und Ingenieure Modelle, um technische Systeme zu realisieren. Sie sind darauf bedacht, Fähigkeiten und Verhalten von Menschen technisch nachzuahmen sowie menschliche Bedürfnisse nachzuvollziehen und zu erfüllen. Das Forschungsfeld der kognitiven Systeme führt die beiden parallelen Entwicklungen in einer Fusion von Geistes-, Natur-, Computer- und Ingenieurwissenschaften zusammen. Es liefert eine zunehmende mathematische Formalisierung kognitiver Modelle, wodurch Fähigkeiten wie maschinelles Erkennen, Lernen, Kommunizieren, Planen und Entscheiden erheblich verbessert werden können. Das wiederum ermöglicht eine Entwicklung von neuartigen, innovativen, nicht nur ”smarten”, sondern kognitiven technischen Systemen.

Die sechs Beiträge des Workshops spannen einen weiten Bogen über wichtige Teilgebiete dieses Forschungsbereichs. Der Schwerpunkt liegt auf aktuellen Entwicklungen und der Verknüpfung von Theorie und Praxis aber auch auf der Vorstellung von neuen Ideen und Ansätzen. Mit dem Workshop sollen sowohl Impulse für die weitere Forschungsarbeit entstehen als auch offene wissenschaftliche Diskussionen zu fachübergreifenden Themen ermöglicht werden.

Organisation. Douglas. W. Cunningham, Petra Hofstedt, Ronald Römer, Ingo Schmid (alle Brandenburgische Technische Universität Cottbus - Senftenberg)

Programmkomitee. Michael Breuß, Christian Hentschel, Petra Hofstedt, Christer Petersen, Ronald Römer, Ingo Schmitt, Matthias Wolff, Thomas Zoglauer (alle Brandenburgische Technische Universität Cottbus - Senftenberg)

[1] BTU Cottbus-Senftenberg, Institute for Informatics, Konrad-Wachsmann-Alle 5, Building VG 1.27, 03046 Cottbus, Germany, douglas.cunningham@b-tu.de
[2] BTU Cottbus-Senftenberg, Communications Engineering, Konrad-Wachsmann-Alle 5, Building 3a, 03046 Cottbus, Germany, ronald.roemer@b-tu.de

Partizipative Gestaltung verlässlicher Assistenzsysteme für die Pflege alter und hochbetagter Menschen

Karsten Weber[1], Alena Wackerbarth[2]

Abstract: Ambient Assisted Living- (AAL) bzw. altersgerechte Assistenzsysteme sollen in der Pflege alter und hochbetagter Menschen, aber auch anderer physisch oder psychisch gehandicapter Personen, eingesetzt werden. Dafür müssen diese Systeme hochverlässlich sein, aber gleichzeitig auch Erwartungen erfüllen, die weniger im technischen und stärker im sozialen Bereich angesiedelt sind. Nach einer kurzen Einführung der Geschichte dieser Systeme und der Beschreibung des Stands der Technik soll auf verschiedene Evaluationswerkzeuge eingegangen werden, die dazu genutzt werden können, eine partizipative Technikgestaltung von AAL-Systemen zu unterstützen.

Keywords: Ambient Assisted Living, altersgerechte Assistenzsysteme, verlässliche Systeme, partizipative Technikgestaltung.

1 Partizipative Technikgestaltung

Altersgerechte Assistenzsysteme zu entwickeln bringt neben technischen Herausforderungen weitreichende soziale Herausforderungen mit sich, da diese Systeme zuhause bei Personen, deren Leben in der Regel durch altersbedingte Handicaps oder Krankheiten gekennzeichnet ist, zu Anwendung kommen sollen. Schon intuitiv ist jedoch leicht zu begreifen, dass die eigene Wohnung oder das eigene Haus nicht nur als physischer Raum angesehen werden dürfen, sondern ein Heim darstellen, das mit Erinnerungen und Erfahrungen gesättigt ist, die gefährdet sein können, wenn sich Gestalt und Charakter dieses Heimes durch (massiven) Technikeinsatz wandeln. Die diesbezügliche Bewertung von Technik in der Pflege- und Gesundheitsversorgung ist jedoch erschwert, da sich viele altersgerechte Assistenzsysteme in der Entwicklung befinden oder erst kürzlich auf den Markt gekommen sind, so dass Erfahrungen nur in unzureichendem Umfang vorliegen. Allerdings bedeutet dieser Mangel nicht, dass eine ethische Evaluation dadurch gänzlich unmöglich würde. Allerdings erfordert die Umsetzung der Ergebnisse einer Evaluation die Einbeziehung der Empirie, so dass entsprechende Ansätze auf soziologische Instrumente zur Befragung von Nutzern existierender Prototypen oder von potenziellen Anwendern in Form von quantitativen und/oder qualitativen Methoden zurückgreifen.

Die vorgestellten Ansätze ethischer Evaluierung und/oder partizipativer Technikentwicklung unterscheiden sich erheblich: Die VDI-Richtlinie 3780 ([Vd91a], [Vd91b])

[1] Lehrstuhl für Allgemeine Technikwissenschaften, Brandenburgische Technische Universität Cottbus, Erich Weinert Str. 1, LG 10, Raum 114, 03044 Cottbus, Karsten.Weber@b-tu.de
[2] Institut für Sozialforschung und Technikfolgenabschätzung, Ostbayerische Technische Hochschule Regensburg, Galgenbergstraße 24, Raum 1.066, 93053 Regensburg, Alena.Wackerbarth@oth-regensburg.de

zielt auf Technik allgemein, MAST ([KRE10], [KEK12]) ausdrücklich nur auf telemedizinische Anwendungen, sentha [DPB07] wiederum auf seniorengerechte Gestaltung von Technik im Allgemeinen, MEESTAR [MWR13] dient der Evaluation insbesondere von AAL-Systemen für das eigene Heim. Darüber hinaus wird nicht in allen Fällen eine normative Beurteilung angestrebt: MEESTAR ist ausdrücklich darauf ausgelegt, sentha behandelt normative Aspekte allenfalls implizit, für MAST sind eher randständig (vgl. [MWF12]), die VDI-Richtlinie 3780 soll ausdrücklich eine normative Technikbewertung ermöglichen, doch wird keine klare Methodik dafür bereitgestellt. Trotz der genannten Unterschiede bleibt festzuhalten, dass es entsprechende Werkzeuge zur ethischen Evaluierung und/oder partizipativer Technikentwicklung gibt; inzwischen fordert das Bundesministerium für Bildung und Forschung (BMBF) bei allen Förderschwerpunkten im Bereich assistiver Systeme die Nutzung solcher Werkzeuge im Entwicklungsprozess.

Die Nutzung partizipativer Verfahren der Technikgestaltung unter Beteiligung möglichst aller Stakeholder ist nicht nur eine ethische Forderung im Sinne des *Empowerment*, sondern auch eine Methode zur Sicherstellung der *Gebrauchstauglichkeit* altersgerechter Assistenzsysteme. Inzwischen findet die Forderung nach Nutzung entsprechender Werkzeuge zunehmend Gehör: Das BMBF auf nationaler Ebene ebenso wie die EU schenkt der Berücksichtigung von ethischen, juristischen und sozialen Aspekten der Technik immer mehr Aufmerksamkeit. Gerade bei verlässlichen Systemen sollte dies eine Selbstverständlichkeit sein.

Literaturverzeichnis

[DPB07] Dienel, H.-L.; Peine, A.; von Blanckenburg, Chr.; Cameron, H.: Die sentha-Methode für die Konzeption seniorengerechter Produkte. In (Friesdorf, W.; Heine, A., Hrsg.): sentha – seniorengerechte Technik im häuslichen Alltag. Ein Forschungsbericht mit integriertem Roman, Springer, Berlin, Heidelberg, 2007; S. 115-139.

[KEK12] Kidholm, Kr.; Ekeland, A. G.; Kvistgaard Jensen, L.; Rasmussen, J.; Duedal Pedersen, Cl.; Bowes, A.; Flottorp, S. A.; Bech, M.: A Model for Asessment of Telemedicine Applications: MAST. International Journal of Technology Assessment in Health Care, 28 (1) 2012; S. 44-51.

[KRE10]Kidholm, Kr.; Rasmussen, J.; Ekeland, A. G.; Bowes, A.; Flottorp, S. A.; Duedal Pederson, Cl.; Kvistgaard Jensen, L.; Dyrehauge, S.: MethoTelemed. Final Study Report, V2.11, 2010, http://www.mast-model.info/Downloads/MethoTelemed_final_report_v2_11.pdf, zuletzt zugegriffen am 30.05.2015.

[KTT06] Kidd, C. D.; Taggart, W.; Turkle, S.: A sociable robot to encourage social interaction among the elderly. Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Orlando, 2006; S. 3972-3976.

[MWF12] Manzeschke, A.; Weber, K.; Fangerau, H.; Rother, E.; Quack, Fr.; Dengler, K.; Bittner, U.: An ethical evaluation of telemedicine applications must consider four major aspects – A comment on Kidholm et al. International Journal of Technology Assessment in

Health Care, 29 (1) 2012; S. 110-111.

[MWR13] Manzeschke, A.; Weber, K.; Rother, E.; Fangerau, H.: Ethische Fragen im Bereich Altersgerechter Assistenzsysteme, VDI/VDE, Berlin, 2013.

[Vd91a] VDI: VDI-Richtlinie 3780. Technikbewertung – Begriffe und Grundlagen, VDI-Verlag, Düsseldorf, 1991.

[Vd91b] VDI: Technikbewertung – Begriffe und Grundlagen. Erläuterungen und Hinweise zur VDI-Richtlinie 3780, VDI-Verlag, Düsseldorf, 1991.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Cognitive Systems: Goals, Approaches, Applications

Ute Schmid[1]

Abstract: In this survey, cognitive systems are characterized in contrast to standard approaches of artificial intelligence and to psychological approaches of cognitive modeling. Arguments for the relevance of symbolic approaches for the development of cognitive systems are presented. After a short overview of current cognitive systems and their applications, the system IGOR will be introduced. IGOR addresses the problem of knowledge level learning, specifically, the inductive acquisition of productive rules from example experience.

Keywords: Cognitive systems, knowledge level, inductive rule learning

A Characterization of Cognitive Systems. Cognitive systems research is of interest not only in the context of artificial intelligence research. Making systems more cognitive can facilitate human interaction with complex software by adapting software to humans instead of making humans adapt to complex software systems [FH06]. For a natural or artificial system to be called cognitive, it typically has to have the ability of high-level cognition, such as reasoning and planning, based on structured knowledge representation rather than simple features [La12]. As research in standard artificial intelligence (AI), most problem domains of interest can only be tackled using heuristic approaches.

In contrast to standard AI, the development of cognitive systems has to be related to models and findings of human cognition and it has to be conducted on system level rather than providing a tool box of unrelated techniques. However, in contrast to the cognitive modeling method of cognitive science, the relation should be qualitative and on the level of general principles rather than quantitative, on the level of detailed comparison of human performance and simulation in a narrow empirical setting. In consequence, research in cognitive systems often needs to be exploratory and cannot only depend on the comparison of methods to measure progress [La12].

Many aspects of cognitive systems are to complex to be described or modelled on the level of neural processing [FP88]. Furthermore, a crucial characteristic of human cognition is the ability to reason over explicit representations and to verbalize and communicate results of such reasoning processes [SK11]. In consequence, statistical models as well as standard black-box machine learning are not sufficient for cognitive systems. While it is indisputable that motor learning, attention and perception are fundamental for higher level cognition, to fulfill the aim to generate software which adapts to humans, the focus of cognitive systems research has to be on the knowledge level.

[1] Universität Bamberg, Fakultät Wirtschaftsinformatik und Angewandte Informatik, Kognitive Systeme, An der Weberei 5, 96045 Bamberg, ute.schmid@uni-bamberg.de

Current Cognitive Systems. Early AI research had its focus on the identification and simulation of general principles of intelligent behavior [NS63]. Over the years, research more and more focused on specialized approaches in different areas such as machine learning or automated planning. However, during the last decade, there is a growing interest in cognitive system research, also under the labels of human-level AI and artificial general intelligence. Two prominent systems are Forbus's Companion Cognitive Systems [FH06] and Langley's Icarus architecture [La06]. The companion system is based on concepts of analogical reasoning and applied as tutor for qualitative physics and as game agent. The Icarus architecture addresses the problem of grounding higher cognition in perception and action and was applied to driver simulation in urban environments. Another cognitive system is IGOR [SK11] which addresses the learning of productive rule sets from example experience. The human ability to generalize over observed regularities is prominent in language learning (e.g., grammar rules) but also in learning relations (e.g., ancestor) and generalized problem solving routines (e.g., Tower of Hanoi). However, machine learning as well as learning in cognitive systems is often restricted to learning flat concepts or reinforcements of observation-action patterns. IGOR is based on an inductive programming approach which identifies regularities in symbolic patterns and provides a recursive generalization. It was successfully applied to learning in problem solving, learning recursive relations and learning simple grammars and it also can solve number series problems as included in many intelligence tests.

Conclusions. When we compare the performance of current cognitive systems with specialized standard AI systems, most current systems cannot compete. However, to advance the goal to make future software adapt to humans, to assist them and to even help to extend human cognition, it will be crucial to focus on the development of systems which integrate different abilities and skills and which rely on principles similar to those on which human cognition is based.

References

[FH06] Forbus, Kenneth D.; Hinrichs, Thomas R.: Companion cognitive systems – a step toward human-level AI. AI Magazine, 27(2):83–95, 2006.

[FP88] Fodor, J. A.; Pylyshyn, Z. W.: Connectionism and cognitive architecture: A critical analysis. Cognition, 1988.

[La06] Langley, P.: Cognitive Architectures and General Intelligent Systems. AI Magazine, 27(2):33–44, 2006.

[La12] Langley, Pat: The cognitive systems paradigm. Advances in Cognitive Systems, 1:3–13, 2012.

[NS63] Newell, A.; Simon, H.A.: GPS, A program that simulates human thought. In (Feigenbaum, E.A.; Feldman, J., eds): Computers and Thought, pp. 279–293. McGraw-Hill, New York, 1963.

[SK11] Schmid, Ute; Kitzelmann, Emanuel: Inductive Rule Learning on the Knowledge Level. Cognitive Systems Research, 12(3):237–248, 2011.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Scale Spaces for Cognitive Systems: A Position Paper

Michael Breuß[1], Douglas W. Cunningham[2] and Martin Welk[3]

Keywords: Scale spaces, partial differential equations, cognitive systems, perception

1 Scale Spaces and Cognitive Systems

Images, like the real world, contain information at many different spatial (and temporal) scales. In image processing and computer vision, the *scale space* is a classic approach for establishing an explicit representation of the multi-scale nature of image content. Introduced by Iijima in 1962 [Ii62] and extensively studied then in Japan [WII99], it has received attention in western countries since Witkin's work [Wi83]. Scale-space theory is an important building block for models of *cognitive systems* as it puts fundamental concepts of the perceptual acquisition of knowledge into a mathematical framework [Ma82, Li13].

The Idea Behind Scale Spaces. Relevant image features (such as the edge of a forest, the trunk of a tree, or the edges of a leaf) in any given image of a scene can be found over a certain – a priori unknown – range of spatial scales. Thus, an image description that emphasizes the hierarchy of scales will be very useful in understanding image content. Basically, an image f is represented as an ordered family of increasingly smoothed versions of the original image. Initially, details observable only at a fine scale (e.g., distant leaves) are removed while large-scale structures are kept. As the amount of smoothing is continuously increased, more and more detail will be removed. Thus, important image structures are ordered by their spatial scale.

Let us denote now the family of scale spaces for an input image f by $\{T_t f : t \geq 0\}$, where the parameter t refers to the scale. Then the scale-space requirements that reflect the need of a hierarchical image representation can be described via three classes of axioms [Al93]: *(i)* axioms related to the underlying *structure of the scale space* such as e.g. the semi-group property $T_{t+s}f = T_t(T_s f)\ \forall t,s \geq 0$; *(ii)* axioms describing *simplification properties* such as e.g. the causality principle which states that structures found at some scale $t^* > 0$ can always be traced back to finer scales $0 \leq t < t^*$; *(iii)* axioms describing *invariances* such as e.g. translation or rotation invariance.
The oldest and best-understood scale space [Ii62] is the *Gaussian scale space*. It is generated by the convolution of an input image with Gaussian kernels K_σ of increasing variance

[1] BTU Cottbus-Senftenberg, Institute for Mathematics, Platz der Deutschen Einheit 1, Building HG 2.51, 03046 Cottbus, Germany, breuss@b-tu.de
[2] BTU Cottbus-Senftenberg, Institute for Informatics, Konrad-Wachsmann-Alle 5, Building VG1c 1.27, 03046 Cottbus, Germany, douglas.cunningham@b-tu.de
[3] UMIT Hall/Tyrol, Institute for Biomedical Image Analysis, Eduard-Wallnöfer-Zentrum 1, 6060 Hall/Tyrol, Austria, martin.welk@umit.at

σ, i.e. $T_t f := K_{\sqrt{2t}} * f$. Equivalently, it can be stated by a solution of the *partial differential equation* (PDE) of linear diffusion $u_t = \Delta u$ where Δ is the Laplace operator, with the input image as initial condition. Here, the scale space parameter t can be interpreted as evolution time. This description immediately gives rise to a variety of generalizations that are based on different PDEs and thus capitalize on the high flexibility of PDEs as a modelling tool. Examples are non-linear diffusion scale spaces [We98] as well as so-called morphological scale spaces [VDBS94]. Since PDEs arise also as Euler-Lagrange equations, this links scale spaces directly to variational methods in continuous-scale optimisation [SW00].

Relation to Visual Perception and Cognitive Systems. A large body of research is dedicated to examining the relationship between mathmatical aspects of scale spaces and aspects of the human visual system (see, e.g., [tHR03, Li94]). For example, Koenderink has shown that the early stages of the visual cortex seem to use a form of multiscale differential geometry to represent images [Ko90]. Likewise, the structure of receptive fields in the retina and of the early visual cortex can be modelled using a Gaussian scale-space. Moreover, the precise layout of receptive field sizes on the retina is exactly what is needed to create a scale-space representation. More intriguingly, the usually ignored fact that there are many more connections feeding from higher visual areas down into lower visual areas than feed upwards from lower to higher areas is consistent with the needs of scale-space construction [tHR03], and has made several predictions (e.g., the temporal modification of receptive field structure) that have since been verified [AB06].

Directions for Future Research. Recent interest in scale-space research goes in numerous directions, see e.g. the conference volume [JFA15]. One field of investigation is the integration of different color models related to human color perception. Another trend is the extension of scale-space models by non-local concepts that allow to exploit similarities between structures occurring at different locations in an image. Other works aim at incorporating representations of texture information into scale-space frameworks. On the application side, there are new attempts to the analysis of audio signals by scale-space methods [LF15].

We believe that the potential of the scale space viewpoint has still not been exhausted. It can be considered a pivotal concept for the theoretical understanding of different approaches to signal and image processing. Its ability to provide mathematical stringent models for processes of human perception together with the versatility of the underlying mathematical concepts lets expect that it can be used for even more detailed modelling of cognitive processes. On this basis, it can guide the development of human-computer interaction processes.

Literatur

[AB06] Angelucci, A.; Bressloff, P. C.: Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. Progress in Brain Research, 154:93–120, 2006.

[Al93] Alvarez, Luis; Guichard, Frédéric; Lions, Pierre-Louis; Morel, Jean-Michel: Axioms and fundamental equations of image processing. Archive for rational mechanics and analysis, 123(3):199–257, 1993.

[Ii62] Iijima, Taizo: Basic theory on normalization of a pattern (in case of typical one-dimensional pattern). Bulletin of the Electrotechnical Laboratory, 26:368–388, 1962.

[JFA15] J.-F.Ãujol, M.Ñikolova, N.P̃apadakis, Hrsg. Scale Space and Variational Methods in Computer Vision. Springer, Cham, 2015.

[Ko90] Koenderink, Jan J.: The brain a geometry engine. Psychological Research, 52(2–3):122–127, 1990.

[LF15] Lindeberg, Tony; Friberg, Anders: Scale-space theory for auditory signals. In: Scale Space and Variational Methods in Computer Vision, S. 3–15. Springer, 2015.

[Li94] Lindeberg, Tony: Scale-space theory in computer vision. Kluwer, Boston, 1994.

[Li13] Lindeberg, Tony: A computational theory of visual receptive fields. Biological cybernetics, 107(6):589–635, 2013.

[Ma82] Marr, D.: Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. MIT Press, 1982.

[SW00] Scherzer, Otmar; Weickert, Joachim: Relations between regularization and diffusion filtering. Journal of Mathematical Imaging and Vision, 12(1):43–63, 2000.

[tHR03] ter Haar Romeny, B.M.: Front end vision and multi-scale image analysis. Kluwer, Dordrecht, 2003.

[VDBS94] Van Den Boomgaard, Rein; Smeulders, Arnold: The morphological structure of images: The differential equations of morphological scale-space. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 16(11):1101–1113, 1994.

[We98] Weickert, Joachim: Anisotropic diffusion in image processing, Jgg. 1. Teubner Stuttgart, 1998.

[Wi83] Witkin, Andrew P.: Scale-Space Filtering. In: Proceedings of the 8th International Joint Conference on Artificial Intelligence. S. 1019–1022, 1983.

[WII99] Weickert, Joachim; Ishikawa, Seiji; Imiya, Atsushi: Linear scale-space has first been proposed in Japan. Journal of Mathematical Imaging and Vision, 10(3):237–252, 1999.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Mathematical structures for modeling semantics, uncertainty, and cognitive processes

Günther Wirsching[1]

Abstract: This talk investigates mathematical structures which are useful for designing cognitive machines with special emphasis on their mathematical properties.

Semantics. Starting point is modeling *semantics* by *order-deterministic pomsets* as defined in [Re96]. It is clear that any syntax tree and any feature structure can be represented by a pomset, hence this model is appropriate for 'literal semantics' as considered in [CEEJKL13, Kapitel 3]. On the other hand, in [WHKLR12] it is suggested that more basic 'pragmatic semantics' as considered in [Sk57, p. 14] and [WL13] are describable by feature-value-relations, and hence also by pomsets.

The *prefix relation* is a partial order on the class of pomsets, or on any set of pomsets over a fixed common alphabet. When we restrict attention to order-deterministic pomsets, then, as is shown in [Re96], the prefix relation has infima and finite suprema. If we have two feature structures which admit *unification* [CEEJKL13, Abschnitt 2.3.3], we obtain this unification by regarding the feature structures as order-deterministic pomsets and taking their supremum.

Uncertainty. For modeling *uncertainty*, consider valuation maps $\Phi : \mathbb{U} \to \mathbb{X}$ defined on a given (finite) set $\mathbb{U}$ of *semantic units*, and associating a *valuation* $\Phi(u) \in \mathbb{X}$ de each semantic unit $u \in \mathbb{U}$ (cf. [WL14]). The valuation $\Phi(u)$ represents uncertainty: when $\mathbb{X} = [0,1]$, an element $x \in \mathbb{X}$ could be interpreted as a probability, or as a fuzzy value. In the present abstract setting, the set $\mathbb{X}$ is not fixed, and other interpretations are also conceivable.

Decision. Now consider also *partial valuation maps* $\Phi : \mathbb{V} \to \mathbb{X}$, where $\mathbb{V} \subseteq \mathbb{U}$ is subset. Denote the domain of definition of a (partial) valuation map by $\mathbb{D}(\Phi)$. Then *decision* is an operation restricting a given (partial) valuation map Φ to some subset

$$\mathbb{D}(\nabla\Phi) \subseteq \mathbb{D}(\Phi) \subseteq \mathbb{U},$$

leading to a partial valuation $\nabla\Phi := \Phi|_{\mathbb{D}(\nabla\Phi)}$ which only valuates the semantic units corresponding to the decision. The elements of $\mathbb{D}(\nabla\Phi)$ are called *winners* of the decision. A decision operation is called *consistent with valuation* [WL14], if there is a binary relation $\rhd$ on $\mathbb{X}$ such that

$$u \in \mathbb{D}(\nabla\Phi) \quad \Leftrightarrow \quad u \in \mathbb{D}(\Phi) \ \wedge\ \forall v \in \mathbb{D}(\Phi) :\ \Phi(u) \rhd \Phi(v). \tag{1}$$

It is proved in [WL14] that, in this situation, properties of ∇ correspond to properties of $\rhd$ as follows:

Decision ∇	**Binary relation $\rhd$**
– non-vanishing	– finite subset-topped
– effective	– antisymmetric
– subset-stable	– transitive

[1] Katholische Universität Eichstätt-Ingolstadt, Mathematisch-Geographische Fakultät, D-85071 Eichstätt, guenther.wirsching@ku.de

This implies that a decision operation ∇ which is consistent with valuation is non-vanishing, effective, and subset-stable, if and only if it is based on a total order $\rhd$ on $\mathbb{X}$ via (1). On the other hand, there is no apparent connection between $\rhd$ and the prefix relation on $\mathbb{U}$.

Join. The second cognitive operation considered here is 'join' of semantic units, and what this means for valuation. To fix notation, let $\Phi_1 : \mathbb{U}_1 \to \mathbb{X}$ and $\Phi_2 : \mathbb{U}_2 \to \mathbb{X}$ be two given valuation maps. For semantic units $u_1 \in \mathbb{U}_1$ and $u_2 \in \mathbb{U}_2$, given as order-deterministic pomsets, denote by $u_1 \sqcup u_2$ their *join* [Re96], or, equivalently, their supremum w.r.t. prefix relation. It may happen that some of the joins $u_1 \sqcup u_2$, when u_1 runs over $\mathbb{U}_1$ and u_2 runs over $\mathbb{U}_2$, do not refer to real situations and therefor can be discarded. The remaining 'meaningful' joins are collected in a subset $\mathbb{V} \subseteq \mathbb{U}_1 \sqcup \mathbb{U}_2$. Moreover, as $\sqcup$ is idempotent, it may happen that $\varnothing \neq \mathbb{U}_1 \cap \mathbb{U}_2 \subseteq \mathbb{V}$. Now a valuation map $\Phi : \mathbb{V} \cup \mathbb{U}_1 \cup \mathbb{U}_2 \to \mathbb{X}$ which extends both Φ_1 and Φ_2 is called *consistent with join*, if there is a binary operation $\vee$ on $\mathbb{X}$ such that

$$\forall u_1 \in \mathbb{U}_1 \text{ and } u_2 \in \mathbb{U}_2 : \quad u_1 \sqcup u_2 \in \mathbb{V} \;\Rightarrow\; \Phi(u_1 \sqcup u_2) = \Phi(u_1) \vee \Phi(u_2).$$

If we assume that $\Phi(u_1 \sqcup u_2) = \Phi(u_1) \vee \Phi(u_2)$ is valid for arbitrary order-determinsitic pomsets u_1, u_2, this would imply that $\vee$ is taking the supremum w.r.t. some partial order $\leqslant$ on $\mathbb{X}$, and that Φ respects the prefix relation in the sense that $\Phi(v) \leqslant \Phi(u)$ whenever v is a prefix of u. Note that at this stage, there is no apparent connection to the total order $\rhd$ considered in (1).

Conclusion. Suppose that we are to design a cognitive system, and we decided to use for modeling semantic units elements from the set $\mathbf{DPOM}(\mathbf{E})$ of order-deterministic pomsets over a given alphabet $\mathbf{E}$. If we wish to model uncertainty by a valuation map $\Phi : \mathbf{DPOM}(\mathbf{E}) \to \mathbb{X}$ which behaves well w.r.t. rational cognitive operations, then it suffices to ensure that Φ maps the prefix relation to a total order on $\mathbb{X}$, and the join $\sqcup$ to taking the maximum w.r.t. that total order.

Keywords: Pomset, order-deterministic pomset, prefix relation, distributive lattice, unification, feature structure, uncertainty, semantic unit, valuation map, decision operation, binary relation, join operation, binary operation.

References

[Re96] Rensink, A.: Algebra and Theory of Order-Deterministic Pomsets. Notre Dame Journal of Formal Logic AI Magazine, Vol. 37: 283–320, 1999.

[Sk57] Sinner, B.F.: Verbal Behavior. Prentice-Hall, Englewood Cliffs, New Jersey, 1957.

[WHKLR12] Wirsching, G.; Huber, M.; Koelbl, C.; Lorenz, R.; Römer, R.: Semantic Dialogue Modeling. Behavioral Cognitive Systems, Lecture Notes in Computer Science, Vol. 7403: 104–113, 2012.

[WL13] Wirsching, G.; Lorenz, R.: Towards meaning-oriented language modeling. IEEE 4rd International Conference on Cognitive InfoCommunication, CogInfoCom2013: 369–374, 2013.

[CEEJKL13] Carstensen K.W.; Ebert, Ch.; Ebert, C.; Jekat, S.; Klabunde, R.; Langer, H.: Computerlinguistik und Sprachtechnologie. Spektrum Akademischer Verlag, Heidelberg 2010.

[WL14] Wirsching, G.; Lorenz, R.: Some algebraic aspects of semantic uncertainty and cognitive biases. IEEE 5rd International Conference on Cognitive InfoCommunication, CogInfoCom 2014.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Perceptual Graphics: The symbiosis between and formal similarity of perceptual psychology and computer graphics.

Douglas W. Cunningham[1]

The environment of each person is overflowing with information, much of which is relevant to some aspect of our perception or behavior. The field of perceptual psychology has spent over 150 years investigating how different sensory systems extract, process, and represent such information. One primary difficulty in this work has been the fact that it is not possible – in principle – to view inside the functional work of any perceptual system. We can at best control the input and observe the output of these dynamic, non-linear, adaptive systems. Computer graphics research has focused over the last 50 years on the effective and efficient visual simulation of such information, traditionally emphasizing a high degree of realism. Trying to re-produce all possible information in any given scene requires nearly infinite resources in processing power and memory. Fortunately, not all the physically present information needs to be synthesized, only the perceptually relevant information.

Although, these two fields have developed independently, they address similar fundamental problems and provide complementary methods for their solution. Furthermore, it is becoming increasingly apparent that specific research questions in either discipline cannot be solved without methods from the other. The fusion of these two complementary approaches not only helps to overcome the inherent limits of each approach, but also opens up entire new realms of scientific inquiry. This integrated approach is referred to as Perceptual Graphics.

[1] BTU Cottbus-Senftenberg, Institute for Informatics, Konrad-Wachsmann-Allee 5, Building VG1c 1.27, 03046 Cottbus, Germany, douglas.cunningham@b-tu.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Einsatzmöglichkeiten kognitiver Retrieval-Modelle in Digitalen Bibliotheken

David Zellhöfer[1]

Abstract: Die Unterstützung bei der Suche nach textuellen Informationen stellt eine typische Aufgabe in modernen Digitalen Bibliotheken dar. Aus Nutzersicht sind die hierfür verwendeten Systeme gereift und es haben sich typische Interaktionsmechanismen, wie die Texteingabeaufforderung, welche beinahe jede gerichtete Suche startet, herausgebildet.

Im Vergleich dazu befindet sich die Suche nach multimedialen oder nicht-textuellen Dokumenten aus Nutzersicht und aus der Sicht der Bibliothekare noch in den Kinderschuhen ? auch wenn dieser Bereich in den letzten Jahren zunehmend an Bedeutung gewinnt. Nach wie vor kämpfen hier jedoch gerichtete und explorative Suchansätze um Nutzerakzeptanz, da es Anwendern mitunter schwerfällt, ihr Informationsbedürfnis zu kommunizieren.

Ein weiterer Punkt, der Multimedia Information Retrieval (MMIR) vom traditionellen, Text-basierten Information Retrieval unterscheidet, ist der Umstand, dass Multimediadokumente nicht notwendigerweise mittels des gleichen Datenzugriffsparadigmas gespeichert sind. Dieser Zustand verkompliziert die Suche in solchen Datenbeständen, da das zu verwendende Retrieval-Modell die jeweiligen Datenzugriffsparadigmen unterstützen muss.

Konsequenterweise müssen die Hauptherausforderungen des MMIR, das Retrieval-Subsystem und die Nutzerinteraktion, ganzheitlich angegangen werden. Solch einen ganzheitlichen, theoretischen Blick auf die MMIR/IR-Forschung wirft das Prinzip der Polyrepräsentation (PdP) ? ein Vertreter der kognitiven Retrievalmodelle. Das PdP geht davon aus, dass verschiedene Repräsentationen, die ein Dokument beschreiben, auf unterschiedlichen, kognitiven Prozessen basieren. Diese Vielzahl an Repräsentationen kann in einer konjunktiven, kognitiven Überlappung zusammengeführt werden, um das aktuelle Informationsbedürfnis des Nutzers zu befriedigen. Hierbei wird angenommen, dass relevante Dokumente in diesem Bereich liegen.

Das PdP als kognitives Modell trifft jedoch keine Aussagen darüber, wie ein entsprechendes Retrieval-Subsystem implementiert werden kann - eine wesentliche Fragestellung der angewandten Informatik. Eine Möglichkeit der Implementierung des PdP stellen Quantenmechanik-basierte IR-Modelle, wie die Commuting Quantum Query Language (CQQL), dar. Die Abfragesprache CQQL ist in diesem Gebiet von besonderer Nützlichkeit, da sie typische Datenzugriffsparadigmen aus den Gebieten DB und IR sowie die unterschiedliche Gewichtung einzelner Repräsentation mit Hinblick auf das aktuelle Informationsbedürfnis unterstützt und außerdem die Personalisierung von Suchergebnissen mittels des Präferenz-basierten Relevance-Feedback-Ansatzes PrefCQQL erlaubt. Mithilfe eines maschinellen Lernverfahrens unterstützt das PrefCQQL-System hierbei den Nutzer bei der Ableitung einer deduktiven Anfrage aus speziellen Präferenzen, die zwischen Beispieldokumenten angegeben werden können. Diese induktiven Präferenzen können genutzt werden, um die Relevanz von Ergebnisdokumenten feingranular zu bewerten. Ferner können induktive Präferenzen bereits zur Anfrageerstellung genutzt werden, so dass sich im Prinzip sämtliche Nutzerinteraktion während der Suche mit PrefCQQL umsetzen lässt.

[1] Staatsbibliothek zu Berlin, Potsdamer Str. 33, 10785 Berlin, david.zellhoefer@sbb.spk-berlin.de

Außerdem lässt sich zeigen, dass CQQL/PrefCQQL verschiedene Strategien der Informationssuche unterstützt, welche typische kognitive Status Suchender bestmöglich unterstützen. Hier seien insbesondere die gerichtete als auch die explorative Suche mittels Facettierung oder Browsing genannt.

Um den vorgestellten, polyrepräsentativen PrefCQQL-Ansatz mit Hinblick auf seine praktische Einsetzbarkeit zu bewerten, wurden verschiedene Experimente durchgeführt. Diese Experimente beantworten primär zwei zentrale Fragen: 1), ob die Hypothesen des PdP im Anwendungsgebiet des MMIR verifizierbar sind und 2), ob ein nutzbares, interaktives MMIR-System auf Grundlagen des PdP und von PrefCQQL implementiert werden kann?

Um die erste Frage beantworten zu können, werden verschiedene Matching-Funktionen, die teilweise dem PdP folgen, mit sechs unterschiedlichen Testdatensätzen, sowohl in einem nicht interaktiven, als auch interaktiven QBE-Szenario gegenübergestellt. Die Ergebnisse dieses Experiments sind uneindeutig.

Im nicht interaktiven Fall zeigt sich keine Überlegenheit der Matching-Funktionen, welche auf dem PdP basieren, gegenüber der Nutzung einzelner Features oder anderen Matching-Funktionen. So übertrifft die Retrieval-Effektivität des arithmetischen Mittels aller Repräsentationen, die der Konjunktion der vorhandenen Repräsentationen, welche der kognitiven Überlappung entspricht. Nichtsdestotrotz zeigt sich, dass Matching-Funktionen, welche dem PdP folgen, stabiler bezüglich ihrer Effektivität sind als Einzelfeatures. Folglich ist die Leistung der PdP-basierten Funktionen verlässlicher. Im interaktiven Szenario, also während der Verwendung von PrefCQQL, können die Voraussagen des PdP verifiziert werden.

Die zweite Frage wird mithilfe eines prototypischen MMIR-Systems, welches als Machbarkeitsstudie eines interaktiven Retrieval-Systems auf Basis von CQQL und PrefCQQL dient, und eines Usability-Tests dessen beantwortet. Des weiteren unterstützt das System unterschiedliche Suchstrategien sowie einen nahtlosen Wechsel zwischen diesen, um Nutzer mit verschiedenen Informationsbedürfnissen bestmöglich zu unterstützen.

Abschließend wird kurz aufgezeigt, welche Möglichkeiten sich aus der Nutzung von kognitiven Retrievalmodellen in Digitalen Bibliotheken ergeben und wie sich die gewonnenen Erkenntnisse in diesen Bereich übertragen lassen. Hier ergeben sich sowohl interessante Möglichkeiten bei der systemübergreifenden, gerichteten Suche in verteilten Nachweissystemen als auch bei der Exploration von Informationsräumen. Durch eine bessere Adressierung der verschiedenen kognitiven Status der Recherchierenden ist damit zu rechnen, dass Nachweissysteme Digitaler Bibliotheken sich vom “digitalen Zettelkatalog” zu wissenserschließenden Informationssystemen entwickeln, welche Nutzer dort abholen, wo sie sich im Rahmen ihrer Forschungs- oder Rechercheaufgaben befinden. Dies stellt eine klare Weiterentwicklung zu aktuell existierenden, verteilten Nachweissystemen dar, welche strenge Anforderungen an das Vorwissen des Nutzers stellen und gleichzeitig nur eine Strategie der Informationssuche unterstützen: die zumeist Bool?sche Schlagwortsuche.

Keywords: Kognitives Information Retrieval, Maschinelles Lernen, Mensch-Maschine-Schnittstelle

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

5. Workshop für Gamification und Virtuelle Welten

5. Workshop für Gamification und Virtuelle Welten

Stefan Stieglitz[1], Christian Scheiner[2], Rüdiger Zarnekow[3], Susanne Robra-Bissantz[4]

Abstract: Der hedonistische Aspekt der Computernutzung gewinnt zunehmend an Bedeutung. Mit der Verbreitung des Internets, unterstützt durch mobile Technologien, sind Computer im letzten Jahrzehnt zum Multimedium geworden und durchdringen bzw. transformieren nahezu alle Bereiche des täglichen Lebens. Unterhaltung, kreative Selbstverwirklichung und freudvolle Interaktion bilden zentrale Aspekte der Nutzung. Das wesentliche Merkmal hedonistischer Informationssysteme besteht darin, dass die Interaktionen mit diesen Systemen sich – zumindest zum Teil – aus sich selbst heraus rechtfertigen. Neben der privaten Nutzung kommen dabei zunehmend auch Anwendungen im Business-Bereich in den Blick. In diesem Zusammenhang wird aktuell insbesondere diskutiert, wie spieltypische Elemente und Prozesse in spielfremden Kontexten nutzbar sind („Gamification") und wie Spiele zur Klärung ernsthafter Fragestellungen beitragen können („Serious Games"). Viele Organisationen sind dabei, die Potenziale des Spielerischen zu entdecken und versprechen sich davon motivationale und kreativitätsfördernde Effekte.

Der Workshop auf der Informatik 2015 soll der Konsolidierung der bisherigen Forschung im deutschsprachigen Raum dienen und einen Ausgangspunkt für die weitere systematische Bearbeitung dieses Themenfeldes markieren. Die einführende Keynote (nicht im Tagungsband enthalten) von Herrn Dr. Philipp Herzig (SAP AG) soll einen praktischen Blick auf das Themenfeld erlauben. Die drei angenommenen Beiträge adressieren verschiedene wichtige Aspekte des eingegrenzten Forschungsfeldes wie den Einsatz von Serious Games im Kontext von Lernen und Wissensvermittlung. Zudem werden virtuelle Welten in einem Beitrag betrachtet..

Ein zentrales Anliegen des Workshops ist es, interessierte Wissenschaftler und Praktiker zusammenzubringen und ein Forum für den weiteren Austausch zu schaffen.

[1] Universität Duisburg-Essen, Fachgebiet Professionelle Kommunikation in elektronischen Medien / Social Media, Forsthausweg 2, 47057 Duisburg, stefan.stieglitz@uni-due.de
[2] Universität zu Lübeck, Institut für Entrepreneurship und Business Development, Ratzeburger Allee 160, 23562 Lübeck, christian.scheiner@uni-luebeck.de
[3] TU Berlin, Fachgebiet Informations- und Kommunikationsmanagement, Straße des 17. Juni 135, 10623 Berlin, ruediger.zarnekow@ikm.tu-berlin.de
[4] TU Braunschweig, Institut für Wirtschaftsinformatik, Mühlenpfordtstraße 23, 38106 Braunschweig, s.robra-bissantz@tu-bs.de

Organisation: Stefan Stieglitz (Universität Duisburg-Essen), Christian Scheiner (Universität zu Lübeck), Rüdiger Zarnekow (TU Berlin), Susanne Robra-Bissantz (TU Braunschweig)

Programmkomitee: Prof. Dr. Christoph Lattemann (Jacobs University Bremen), Prof. Dr. Thomas Knieper (Universität Passau), Prof. Dr. Rolf Nohr (Hochschule für Bildende Künste Braunschweig), Prof. Eku Wand (Hochschule für bildende Künste Braunschweig), Prof. Dr. Ulrike Lechner (Universität der Bundeswehr München), Prof. Dr. Susanne Strahringer (TU Dresden), Dr. Maximilian Witt (TU Braunschweig, Institut für Wirtschaftsinformatik), Dr. Dr. Albrecht Fritzsche (Universität Erlangen-Nürnberg)

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Micropayments in virtuellen Welten – Prozessmodell und Nutzung bei 7- bis 12-jährigen Besuchern

Andreas Mladenow[1], Niina Maarit Novak[1], Christine Strauss[1]

Abstract: Durch verstärkte Nutzung mobiler Endgeräte bei jüngeren Zielgruppen erleben auf virtuellen Welten zugeschnittene Geschäftsmodelle derzeit einen Aufschwung. Diese Modelle beruhen auf Umsätzen durch Kleinstbeträge für den Erwerb virtueller Güter und „follow the free"-Mitgliedschaften. Doch wie steht es um die Akzeptanz bei den Mitgliedern dieser Zielgruppe? Der vorliegende Beitrag analysiert die vorzufindenden Prozessmodelle von Micropayments in virtuellen Welten und stellt ein daraus resultierendes, generisches Prozessmodell im Detail vor. Ferner wird die Nutzerakzeptanz auf Basis einer Befragung unter 200 Kindern anhand getätigter Ausgaben in virtuellen Welten untersucht und interpretiert. Dabei werden einerseits über die Höhe der Ausgaben, sowie über die eingesetzten mobilen Geräte Aussagen getroffen, andererseits werden die Resultate der Befragung 2013 und 2014 verglichen.

Keywords: Virtuelle Welt, Virtuelle Kinderwelt, Virtuelle Güter, Micropayments, Prozessmodell, Geschäftsmodell, Business Model, Digital Goods, Intangible Goods, Digitale Güter, Gaming Industry, Communities, Digital Natives, junge Zielgruppe, Nutzerakzeptanz, Ubiquitous Computing

1 Einleitung

In den vergangenen Jahren hat die Nutzung mobiler Endgeräte insgesamt stark zugenommen, dies gilt auch für die Nutzung mobiler Endgeräte durch Kinder. Dabei stellt sich die Frage, wie Kinder mit neuen Technologien umgehen, wie sie Internetapplikationen anwenden, und wie sie die Ubiquität der Nutzung durch mobile Endgeräte für sich umsetzen [Be12, Ml14]. In diesem Zusammenhang erfreuen sich virtuelle Welten gerade bei den jüngeren Anwendern großer Beliebtheit [Ma10, No13], während sich virtuelle Welten in unternehmerischem Kontext nicht immer als erfolgreich erwiesen haben [Yo13].

Um Einnahmen zu generieren, setzen Betreiber und Manager von virtuellen Welten heute zunehmend auf Zahlung sogenannter Micropayment-Transaktionen. Dabei kann der User virtuelle Gebrauchsgegenstände mittels Klein- bzw. Kleinstzahlungen erwerben. Doch welche Akzeptanz haben solche Micropayment-Transaktionen bei der jungen Zielgruppe [Ma10]? Welches Prozessmodel findet in Bezug auf Micropayments statt?

[1] Universität Wien, Fakultät für Wirtschaftswissenschaften, Fachbereich *Electronic Business*, Oskar-Morgenstern-Pl. 1, 1090 Wien, {andreas.mladenow, niina.novak, christine.strauss}@univie.ac.at

Zu diesem Themenkomplex gibt es bisher nur wenige aktuelle wissenschaftliche Beiträge [Ba07, Me09, No14]. Vor diesem Hintergrund untersucht der vorliegende Beitrag derzeit anzutreffende Mikropayment-Transaktionen in virtuellen Welten und analysiert die in diesem Zusammenhang bestehende Nutzerakzeptanz bei 7- bis 12-jährigen Usern basierend auf zwei durchgeführten Befragungen unter rund 200 Kindern im Rahmen von Kinderuni-„Vorlesungen“ an der Universität Wien [Ki15].

Der vorliegende Beitrag ist in vier Kapitel gegliedert. Das folgende Kapitel 2 beinhaltet theoretische Grundlagen zu virtuellen Welten, Motivationstreiber, Potenziale und Akzeptanz seiner User und Micropayments für virtuelle Güter. Kapitel 3 beschreibt das Prozessmodell zu Micropayment-Transaktionen, führt eine Reihe von Praxisbeispielen an und präsentiert die Ergebnisse einer Befragung aus 2013 und 2014 unter 7- bis 12-jährigen Nutzern im Rahmen einer Kinderuni-Lehrveranstaltung. Danach schließt der Beitrag mit einer Conclusio und einem kurzen Ausblick.

2 Theoretische Grundlagen

Virtuelle Welten beruhen auf einer computer-gestützten, simulierten, künstlichen Welt [Ka09]. Basierend auf unterschiedlichen Definitionen in der Literatur unterstreicht Bell [Be08] die folgenden Eigenschaften, die für virtuelle Welten gelten: partizipierende Teilnehmer, synchrone Kommunikation, die Vertretung durch einen Avatar, Persistenz sowie die Vernetzung durch Informations- und Kommunikationstechnologie (IKT). Die simultane Partizipation wird als grundlegende Voraussetzung betrachtet, wobei die Intensität der Interaktionen durch jeweilige Ausprägungen der virtuellen Welt bedingt ist. Diese Ausprägungen können sich gerade in virtuellen Kinderwelten stark von den „irdischen“, realen, physikalischen Gesetzen unterscheiden. Daher kann es sich auch bei den Avataren um Fantasiewesen handeln [Ad08]. Die Einbindung von Avataren und die Computervernetzung unterscheiden virtuelle Welten von einem real-physischen und real-sozialen Umfeld. Da virtuelle Welten nicht „stillstehen“, wenn ein Nutzer sie verlässt, beinhaltet die Persistenz eine weitere Einschränkung, die somit andere Vertreter aus der Spielewelt ausschließen.

Die Frage nach dem Antrieb und der Motivation [Ei12, Ml11, Ml12], weshalb viele Menschen zunehmend Zeit in virtuellen Welten verbringen, wurde in den letzten Jahren eingehend untersucht, wobei zwischen intrinsischen und extrinsischen Faktoren unterschieden wird [Ka10, Ve11, Zh11]. Wirklichkeitsflucht und Unterhaltungsdrang werden als intrinsische, Incentives und Bedienungsfreundlichkeit als extrinsische Faktoren in virtuellen Welten angesehen [Ve11]. In diesem Kontext wird zwischen drei Arten von Motivationstreibern unterschieden [No13], die User dazu veranlassen virtuelle Welten aufzusuchen. Tab. 1 gibt einen Überblick über diese Treiber (funktional, experimentell, sozial) mit den vornehmlichen Partizipationsgründen.

Die speziellen Charakteristika virtueller Welten und die damit verbundenen spielerischen Möglichkeiten der Kommunikation eröffnen gerade dem jüngeren

Zielpublikum einen bereichernden Ort der Begegnung [No14]. Zudem suchen Nutzer eher jene Welten auf, in denen sie auf Peers mit ähnlichen Motivationstreibern und Interessen treffen [Ei12].

Neben den Motivationstreibern für die Nutzung wird in der Literatur auf Chancen und Herausforderungen virtueller Welten eingegangen [Ka10]. Dabei wird vor allem dem Bildungsbereich ein bedeutungsvolles Potenzial eingeräumt [Wa09, Bo07], auf die Nutzung bei der Ausbildung von Pädagogen und Pädagoginnen für Kinder hingewiesen [Di11] sowie auf die Wirkung bei Kindern [OK11]. Während wissenschaftliche Beiträge eher auf Nutzerverhalten bei spielebasierten virtuellen Lernumgebungen fokussieren [Fa13], existieren hinsichtlich Nutzerakzeptanz von virtuellen Welten und die entsprechende Zahlungsbereitschaft bei jüngeren Zielgruppen noch vergleichsweise wenig Untersuchungen [Mä11]. Die theoretischen Grundlagen stützen sich dabei vor allem auf sogenannte Technologieakzeptanzmodelle wie der Unified Theory of Acceptance and Use of Technology von Venkatesh et al. [Ve12].

Unter Micropayment wird ein Zahlungsverfahren geringer Transaktionssummen verstanden [Re10]. Darunter fallen Kleinstbeträge ab € 0.01; die Höchstgrenze wird unterschiedlich angesetzt; sie liegt zumeist bei etwa € 5.00 [Ve13]. Erst mit der zunehmenden Verfügbarkeit des World Wide Web durch Breitband-Internet und der Ubiquität durch mobile Endgeräte erlangen online-basierte Micropayment-Transaktionen zunehmend Relevanz. In virtuellen Welten werden solchen Kleinstzahlungen für den Ankauf von virtuellen (intangiblen) Gütern benutzt, die in der virtuellen Welt Verwendung finden können.

Motivationstreiber	Spielertypus	Motivation der Teilnehmer
funktional	„Bewohner“	vordefinierte Ziele erreichen (shoppen, spielen, Geld verdienen)
	„Schöpfer“	eigene Ziele setzen, sich selbst herausfordern
experimentell	„Erkunder“	den vor allem intrinsischen Drang nach neuen Erlebnissen befriedigen (Spaß, Entertainment, Rollenspiele)
	„Flüchtling“	in eine virtuelle Welt untertauchen, wo Träume und Fantasien ausgelebt werden können
sozial	„Sozialisierer“	Wunsch andere User zu treffen und mit ihnen zu interagieren

Tab. 1: Motivationstreiber für die Nutzung virtueller Welten [No13]

3 Micropayment-Transaktionen in virtuellen Welten

Im Folgenden werden nun Prozessmodell und Praxisbeispiele beschrieben. Ferner wird auf die Befragungen und deren Ergebnisse der 7- bis 12-jährigen User im Rahmen der Erhebungen von 2013 und 2014 eingegangen.

3.1 Prozessmodell

Um ein tragfähiges generisches Prozessmodell erstellen zu können, wurden rund 50 software-basierte virtuelle Welten für PCs und mobile Endgeräte in einer pre-study analysiert. Berücksichtigt wurden dabei kostenlos zugängliche, deutschsprachige Online-Welten mit Client-Server bzw. Peer-to-Peer Architektur für die analysierte Zielgruppe (ab 6 Jahren). Die Analyse ergab, dass ein bestimmtes Prozessmodell basierend auf Prepaid-Micropayments dominiert (Abb. 1).

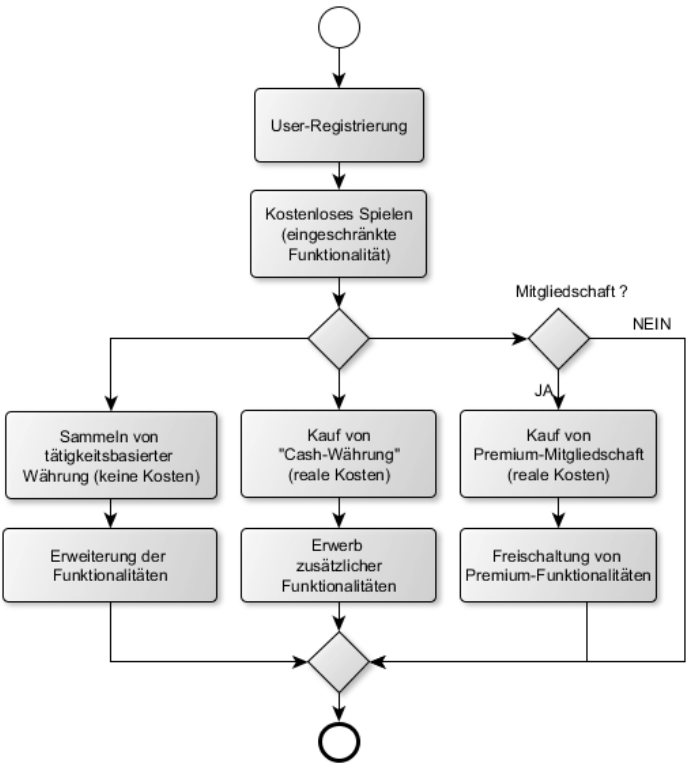


Abb. 1: Micropayment-Prozessmodell bei virtuellen Welten

Zunächst kann sich der Besucher kostenlos registrieren und eingeschränkte Basisfunktionen der virtuellen Welt nutzen. Als Belohnung für seine Aktivitäten erhält er einen Betrag in einer tätigkeitsbasierten Währung. Diese Währung ermöglicht weitere Funktionalitäten in Anspruch zu nehmen und Aktivitäten durchzuführen. Neben dem kostenlosen Spielen hat der Nutzer nun auch die Möglichkeit eine zweite Cash-Währung zu erwerben, die allerdings reale Kosten verursacht. Die Cash-Währung kann dazu verwendet werden, um weitere Funktionalitäten zu kaufen und somit freizuschalten. Ebenso steht dem Besucher die Möglichkeit offen eine Premium-Mitgliedschaft zu erwerben, die typischerweise für einen Zeitraum von einem Monat bis zu einem Jahr angeboten wird und dem Nutzer die Verwendung von Premium-Funktionalitäten erlaubt. Der Nutzer muss also stets zuerst eine Zahlung leisten, ein *pre-payment*, um anschließend neue Funktionalitäten nutzen zu können.

Zur näheren Beschreibung des Prozessmodells wurden drei Praxisbeispiele (Neopets, Bin Weevils und Goodgame Empire) ausgewählt [Ne15, Bi15, Go15] wobei als Auswahlkriterien die Beliebtheit des Spiels im deutschsprachigen Raum (Anzahl der Downloads bzw. registrierte Nutzer und unterschiedliche Geschlechteransprache der Zielgruppe durch den Betreiber) herangezogen wurde.

3.2 Praxisbeispiele

Neopets ist eine virtuelle Welt, in der Spieler virtuelle Haustiere, sogenannte *Neopets*, halten können. Die Besucher dieser virtuellen Welt kümmern sich um das Wohl dieser Tiere und nehmen gemeinsam mit ihnen an Spielen und weiteren sozialen Aktivitäten teil [Ne15].

Neopoints (NP) können durch die Teilnahme an Spielen gesammelt werden, sowie durch den Verkauf von Gegenständen und weiteren Transaktionen. *Neocash* (NC) hingegen kann nur im NC-Center per Kreditkarte, Paypal-Überweisung oder durch Einlösung einer *Neocredits-Karte* (Gutscheinkarte) erworben werden. *Neocash* kann ausschließlich im NC-Shoppingcenter, der *NC-Mall,* verwendet werden. Es gibt vier Kaufoptionen um *Neocash* zu erwerben: 500 NC zu USD 5,00, 1000 NC zu USD 10,00, 3000 NC zu USD 30,00 und 5500 NC zu USD 50,00; zu den genannten Beträgen kommen noch jeweils Steuern von 20% des angegebenen Preises. Darüber hinaus sind *Neocash* und *Neopoints* nicht konvertierbar.

Die Wirtschaft von *Neopets* basiert auf *Neopoints*, der Währung in der virtuellen Welt *Neopets*. *Neopoints* können bei der *neopianischen Nationalbank* auf ein Bankkonto eingezahlt werden um Zinsen zu erhalten. Sparen macht also in der virtuellen Welt durchaus Sinn. *Neopoints* können verwendet werden um Gegenstände von anderen Spielern zu erwerben, um Aktien an der *neopianischen Börse* zu erwerben oder um in einer Vielzahl an Geschäften oder dem *neopianischen Basar* beispielsweise Nahrungsmittel, Spielzeug und sogar Kleidung für die virtuellen Haustiere zu kaufen.

Bei den Premium-Mitgliedschaften, die es unter anderem ermöglichen ein zusätzliches Haustier zu halten, doppelte Neopunkte bei Premium-Spezialspielen zu erlangen und zu monatlichen Bonus-NPs führen, kann aus drei Alternativen ausgewählt werden: Jahresmitgliedschaft zu USD 69,95, viermonatige Mitgliedschaft zu USD 24,94 und einer einmonatigen Mitgliedschaft zu USD 7,95.

Bin Weevils ist eine virtuelle Spielewelt in der Kinder in Form von virtuellen Fantasietieren, sogenannten *Weevils*, an einer Vielzahl von kleinen Online-Spielen teilnehmen können. Ein *Weevil* lebt in der virtuellen Welt namens *Bin,* und verbringt neben der Teilnahme an Spielen auch viel Zeit damit sein Zuhause einzurichten und im angrenzenden Garten Pflanzen und Früchte zu züchten [Bi15].

Mulch ist die virtuelle Währung in *Bin Weevils*. *Weevils* können durch die Teilnahme an Spielen und das Ernten von Früchten und Gemüse *Mulch* verdienen. *Dosh*, die zweite virtuelle Währung, kann jedoch nur von *Bin Tycoons*, den Premium-Mitgliedern in *Bin Weevils*, durch Kredikartenzahlung, Paypal-Überweisung oder durch Einlösung einer Mitgliedschaftskarte erworben werden. Um *Bin Tycoon* zu werden stehen drei verschiedene Mitgliedschaftsoptionen zur Auswahl: Jahresmitgliedschaft zu USD 39,95, halbjährliche Mitgliedschaft zum Preis von USD 29,95, sowie eine einmonatige Mitgliedschaft zu USD 6,95 pro Monat. *Dosh* kann in folgenden Einheiten erworben werden: 10 Dosh zu USD 2,00, 50 Dosh zu USD 5,95, 150 Dosh zu USD 14,95 und 360 Dosh zu USD 29,95.

Stellvertretend für die sogenannte „*Goodgame-Reihe"* (Shadow Kings – Dark Ages, Shadow Kings – Empire, Goodgame Big Farm, Goodgame Galaxy, Goodgame Empire, Goodgame Fashion, Goodgame Disco, Goodgame Café, Goodgame Mafia, Goodgame Poker) an Browserspielen wird hier *Goodgame Empire* betrachtet [Go15]. Das Spiel ist als *Empire-Four Kingdoms* auch als Mobile-App für iOS und Android Mobiltelefone verfügbar.

Goodgame Empire ist eine Mischung aus Strategie-Browserspiel und Städtebauspiel. Die Spieler verfolgen als Burgherren mehrere Ziele. Einerseits geht es darum sich eine Festungsanlage aufzubauen, sich mit Mitspielern in Allianzen zu verbünden, den Angriffen von anderen Festungsinhabern standzuhalten sowie strategische Militäreinsätze durchzuführen; andererseits gilt es ein funktionierendes Wirtschaftssystem aufzubauen, in dem Rohstoffe produziert werden um anschließend damit bauen und Handel betreiben zu können.

Rubine und *Goldmünzen* sind die virtuellen Währungen in *Goodgame Empire*. *Münzen* erhält man für das Eintreiben von Steuern, sowie für Siege (*Quests*) und die erfolgreiche Verteidigung der eigenen Festung. Rubine gibt es ebenfalls für zahlreiche *Quests*, jedoch können Rubine auch zusätzlich im *Paymentshop* für reales Geld erworben werden. Rubine können in verschiedenen Paketen von 2.500 Rubinen zu einem Preis von EUR 1,99 bis 180.000 Rubinen zu einem Preis von EUR 99,00 erworben werden. Darüber hinaus können im Paymentshop auch Holz (15.000 Stück pro Woche), Steine (15.000 Stück pro Woche), Goldmünzen (20.000 Stück pro Woche) und Rubine (5.500 Stück pro

Woche), jeweils für einen Zeitraum von sechs Wochen zum Preis von EUR 19,90 gekauft werden. Die Zahlungsmöglichkeiten sind ebenso vielfältig wie die Produktpalette.

Im folgenden wird ein Micropayment-Prozessmodell beschrieben um einen Einblick in einzelne Sequenzen zu geben; Abbildung 2 zeigt exemplarisch anhand von *Goodgame Empire* die dort implementierten Prozesse und deren Abfolge. Nachdem sich der User registriert hat, kann das Spiel in den ersten fünf Levels ausschließlich kostenfrei gespielt werden. Erst danach wird ein Paymentshop freigeschaltet, und es können Rubine und andere virtuelle Güter gekauft werden. Neben den Standardzahlungsmöglichkeiten (Kreditkarte, Paypal) ist die Bezahlung auch per Sofortbanking, Zahlung per Rechnung, Zahlung mit Festnetz und Mobilfunk, SMS, Amazon-Payments, Maestro, Ukash, EPS, Skrill und JCB/Diners Club möglich. Premium-Mitgliedschaften werden bei *Goodgame Empire* nicht angeboten.

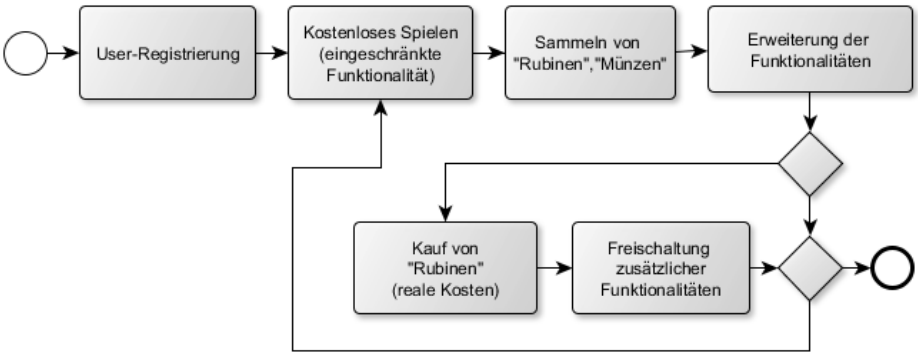


Abb. 2: Micropayment-Prozessmodell bei *Goodgame Empire* (Stand: Mai 2015)

3.3 Nutzerbefragung zu Micropayment-Transaktionen in virtuellen Welten

Die KinderuniWien ist eine jährlich stattfindende Veranstaltungsreihe mehrerer Wiener Universitäten und Fachhochschulen mit insgesamt etwa 4.500 teilnehmenden Kindern im Alter von 7 bis 12 Jahren. In den Jahren 2013 und 2014 wurden im Rahmen der Vorlesung „Was kostet Deine Reise in die virtuelle Kinderwelt?“ die teilnehmenden Kinder zur Nutzung von virtuellen Welten und den damit verbundenen Ausgaben befragt. An der eineinhalbstündigen „Vorlesung“ nahmen jeweils 150 studierende Kinder teil; davon haben insgesamt 101 (im Jahre 2013) bzw. 106 TeilnehmerInnen (2014) den Fragebogen vollständig ausgefüllt.

Eine Gegenüberstellung der Micropayment-Ausgaben von 7- bis 12-jährigen Kindern im Jahr 2013 und im Jahr 2014 ist aus Abb. 3 ersichtlich. Der Anteil an TeilnehmerInnen,

die keinerlei Ausgaben in virtuellen Welten tätigten, hat sich dabei deutlich von über 80% (n=82) im Jahr 2013 auf knapp 55% (n=58) im Jahr 2014 reduziert. Zudem haben sich im Vergleichszeitraum die Prozent-Werte für Ausgaben zwischen EUR 1,00 und EUR 5,00 verdoppelt, für Ausgaben zwischen EUR 5,00 und EUR 20,00 verfünffacht und für Ausgaben über EUR 20,00 verdreifacht. Die *tägliche* Nutzung mobiler Endgeräte hat sich bei den TeilnehmerInnen im selben Zeitraum deutlich erhöht (Smartphones von 28% auf 38%, Tablets von 9% auf 22%; Rundung auf ganze Prozentpunkte), wohingegen *gar keine Nutzung* jeweils bei beiden Endgerätetypen deutlich zurückgegangen ist, nämlich bei Smartphones von 34% auf 23% und bei Tablets von 57% auf 29%.

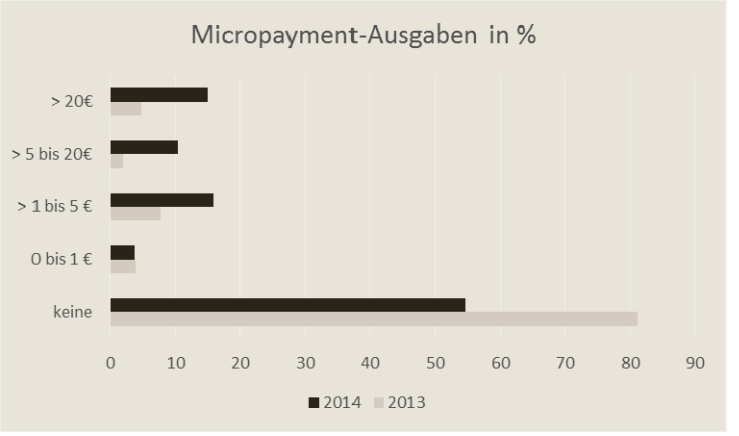


Abb. 3: Vergleich der 2013 und 2014 erhobenen Micropayment-Ausgaben auf Jahresbasis in virtuellen Welten (in %)

4 Zusammenfassung und Ausblick

Die Nutzung virtueller Welten nimmt im Kontext steigender Benutzung und verbesserter Technologie von mobilen Endgeräten gerade bei der jungen Zielgruppe rasch zu. Dabei spezialisieren sich Unternehmen der Gaming-Industrie auf entsprechende micropayment-basierte Geschäftsmodelle in diesem Marktsegment.

Vor diesem Hintergrund untersucht der vorliegende Beitrag anzutreffende Prozessmodelle bei Micropayment-Transaktionen in virtuellen Welten für eine junge Zielgruppe und ermittelt ein dominierendes Prozessmodell. Eine Befragung bei 7- bis 12-jährigen Besuchern von virtuellen Welten, die 2013 und 2014 im Rahmen von Kinderuni-„Vorlesungen“ durchgeführt wurden, zeigte eine Verdopplung der Anzahl der Besucher virtueller Welten innerhalb eines Jahres, die zwischen EUR zu EUR 5,00 pro Jahr ausgeben; hingegen hat sich die Anzahl der Besucher, die zwischen EUR 5,00 und

EUR 20,00 ausgeben sogar verfünffacht und die Anzahl der Besucher, die mehr als EUR 20,00 ausgeben, verdreifacht. Ferner konnte eine deutliche Erhöhung täglicher Nutzung mobiler Endgeräte festgestellt werden. In diesem noch wenig beachteten Forschungsbereich sind weitere Datenerhebungen und -analysen für 2015 und Folgejahre vorgesehen, um weitere Einblicke über diese interessante Entwicklung erlangen zu können.

Literaturverzeichnis

[Ad08] Adrian, A. (2008) No one knows you are a dog: Identity and reputation in virtual worlds. Computer Law & Security Report 24, 366-374.

[Ba07] Bainbridge, W. S. (2007) The scientific research potential of virtual worlds. Science 317 (5837), 472-476.

[Be12] Becker, A., Mladenow, A., Kryvinska, N., Strauss, C. (2012). Evolving taxonomy of business models for mobile service delivery platform. Procedia Computer Science, 10, 650-657.

[Be08] Bell, M. W. (2008). Toward a definition of "virtual worlds". Journal For Virtual Worlds Research, 1(1).

[Bi14] Bin Weevils, www.binweevils.com, Zugriff: 01.05.2015

[Bo07] Boulos, M. N. K., Hetherington, L., Wheeler, S. (2007). Second Life: an overview of the potential of 3-D virtual worlds in medical and health education. Health Information & Libraries Journal, 24(4), 233-245.

[Ca05] Castronova, E. (2005) Synthetic Worlds – The Business and Culture of Online Games. The University of Chicago Press. Chicago, USA.

[Ca05] Castronova, E. (2005) Synthetic Worlds – The Business and Culture of Online Games. The University of Chicago Press. Chicago, USA.

[Di11] Dickey, M. D. (2011). The pragmatics of virtual worlds for K-12 educators: investigating the affordances and constraints of Active Worlds and Second Life with K-12 in-service teachers. Educational technology research and development, 59(1), 1-20.

[Ei12] Eisenbeiss, M., Blechschmidt, B., Backhaus, K., Freund, P. A. (2012) The (Real) World Is Not Enough:" Motivational Drivers and User Behavior in Virtual Worlds. Journal of Interactive Marketing 26, 4-20.

[Fa13] Faiola, A., Newlon, C., Pfaff, M., & Smyslova, O. (2013). Correlating the effects of flow and telepresence in virtual worlds: Enhancing our understanding of user behavior in game-based learning. Computers in Human Behavior, 29(3), 1113-1121.

[Ne14] Goodgame Empire, http://empire.goodgamestudios.com, Zugriff: 03.05.2015

[Ki15] Kinderuni Wien, Was kostet deine Reise in die virtuelle Kinderwelt? http://2013.kinderuni-anmeldung.at/event.php?event_id=561&field_id=1, Zugriff: 15.4.2015

[Ka09] Kaplan, A. M., Haenlein, M. (2009) The fairyland of Second Life: Virtual social worlds and how to use them. Business Horizons 52, 563-572.

[Ka10] Kaplan, A. M., Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business horizons, 53(1), 59-68.

[Ma10] Marsh, J. (2010). Young children's play in online virtual worlds. Journal of Early Childhood Research, 8(1), 23-39.

[Mä11] Mäntymäki, M., Salo, J. (2011). Teenagers in social virtual worlds: Continuous use and purchasing behavior in Habbo Hotel. Computers in Human Behavior, 27(6), 2088-2097.

[Me09] Messinger, P. R., Stroulia, E., Lyons, K., Bone, M., Niu, R. H., Smirnov, K. and Perelgut, S. (2009) Virtual worlds – past, present, and future: New directions in social computing. Decision Support Systems 47, 204-228.

[Ml11] Mladenow, A., Fröschl, K. A. (2011). Kooperative Forschung. Frankfurt am Main: Peter Lang.

[Ml12] Mladenow, A., Kryvinska, N., Strauss, C. (2012). Towards cloud-centric service environments. Journal of Service Science Research, 4(2), 213-234.

[Ml14] Mladenow, A., Bauer, C., Strauss, C. (2014). Social Crowd Integration in New Product Development: Crowdsourcing Communities Nourish the Open Innovation Paradigm. Global Journal of Flexible Systems Management, 15(1), 77-86.

[Ne14] Neopets, www.neopets.com, Zugriff: 02.05.2015

[No13] Novak, N. M., Mladenow, A., Strauss, C. (2013). Avatar-based Innovation Processes – Are Virtual Worlds a breeding ground for Innovations? In Proceedings of International Conference on Information Integration and Web-based Applications & Services (p. 174). ACM.

[No14] Novak, N. M., Mladenow, A., Strauss, C. (2014). Virtual worlds as settings for avatar-based innovation processes. Journal of Service Science Research, 6(1), 71-98.

[OK11] O'Keeffe, G. S., & Clarke-Pearson, K. (2011). The impact of social media on children, adolescents, and families. Pediatrics, 127(4), 800-804.

[Pe11] Pearce, C., Boellstorff, T., Nardi, B. A. (2011). Communities of play: Emergent cultures in multiplayer games and virtual worlds. MIT Press.

[Re10] Reddy, S., Estrin, D., Hansen, M., Srivastava, M. (2010, September). Examining micro-payments for participatory sensing data collections. In Proceedings of the 12th ACM international conference on Ubiquitous computing (pp. 33-36). ACM.

[St15] Statista, http://de.statista.com/statistik/daten/studie/301018/umfrage/tablet-nutzer-in-oesterreich-nach-alter-und-anzahl-der-tablets/, Zugriff: 25.4.2015

[St08] Stieglitz, S. (2008). Steuerung Virtueller Communities: Instrumente, Mechanismen, Wirkungszusammenhänge. Springer-Verlag.

[Ve13] Verme, P. L. H., Benavides, R. A. V. (2013). Virtual currencies, Micropayments and the payments systems: a challenge to fiat money and monetary policy?. European Scientific Journal, 9(19).

[Ve12] Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS quarterly, 36(1), 157-178.

[Wa09] Warburton, S. (2009). Second Life in higher education: Assessing the potential for and the barriers to deploying virtual worlds in learning and teaching. British Journal of Educational Technology, 40(3), 414-426.

[Yo13] Yoon, T. E., George, J. F. (2013) Why aren't organizations adopting virtual worlds? Computers in human Behaviour 29, 772-790.

[Zh11] Zhou, Z., Jin, X., Vogel, D. R., Fang, Y., Chen, X. (2011) Individual motivations and demographic differences in social virtual world uses: An exploratory investigation in Second Life. International Journal of Inf. Management 31, 261-271.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

How Games and Game Elements Facilitate Learning and Motivation: A Literature Review

Christian Karl Grund[1]

Abstract: Games and game elements are increasingly used by organizations to facilitate learning and motivation, often without a clear understanding how they actually achieve these effects. This may lead to the insufficient design and use of serious games as well as gamified applications. Since the development of such applications is usually expensive, high costs and few realized benefits might result. In order to avoid such unfavorable outcomes, it is necessary to understand the underlying mechanisms that lead to learning and motivation. For this purpose, this study reviews research articles describing the use of games and game elements beyond entertainment and outlines their theoretical foundations. Based on the resulting insights, several theory-driven design guidelines for game-based learning and motivation are derived. This study therefore equips researchers with theoretical insights on how games and game elements facilitate learning and motivation, and practitioners with theory-driven design guidelines for their design and evaluation.

Keywords: Serious Games, Gamification, Games with a Purpose, Theoretical Foundation.

1 Introduction

Games and game elements are increasingly utilized by organizations to facilitate learning and motivation [LLS13]. They are employed for example as "serious games" to create experiential learning environments that fulfill more goals than simply entertaining players [Ab87]. Rather, they aim at advancing these players by improving their capabilities or knowledge [ZHR12]. Another possible use of games and game elements is "gamification" [De11]. Here, game elements are used in a non-game context in order to achieve motivational effects [Ka12]. Beyond that, there are games that solve real-world problems just by being played [ALB06]. In the game "Peekaboom" for example, players identify objects in pictures and thus enhance a computer vision algorithm [ALB06].

Despite successful applications already developed, previous research lacks systematic investigations of the mechanisms of gamification [Ka12], or how instructional theories can frame the design of serious games [Ch10]. Without such an understanding, serious games might end up as "drill and practice activities sugar-coated with game characteristics" [Ch10] and gamification might be perceived as "exploitationware" [Bo15] instead of intrinsically motivating participants. Since the development of such applications is usually expensive [Bo10], their insufficient design might result in high costs and few realized benefits. It is therefore necessary to understand how games and game elements

[1] University of Augsburg, Faculty of Business and Economics, Universitätsstr. 16, 86159 Augsburg, Germany, christian.grund@wiwi.uni-augsburg.de

lead to learning and motivation, because this understanding helps to frame the design and evaluation of "gameful" experiences for learning and motivation. In contrast to existing theoretical approaches, that focus mostly on motivation theories, this study provides a holistic view on games and game elements, which results in considering both motivation and learning theories. Hence, this study seeks to provide answers to the following two research questions:

RQ1: How do games and game elements facilitate learning and motivation?

RQ2: How can theory frame the design of game-based learning and motivation?

In order to answer these questions, this study conducts a literature review, as this research method can be used to establish a theoretical foundation for an emerging issue [WW02]. For this purpose, the theoretical foundations of research articles describing the use of games and game elements beyond entertainment are examined. Based on the resulting insights, this study suggests several theory-driven design guidelines for game-based learning and motivation, thus providing an answer to the second research question.

The remainder of this paper is organized as follows: Section 2 defines different ways of using games and game elements beyond entertainment. The method and search setup employed are described in section 3. A review of the theoretical foundation of using games and game elements for learning and motivation is afterwards conducted in section 4, followed by the resulting design guidelines for game-based learning and motivation in section 5. The paper closes in section 6 with a conclusion and possibilities for future research.

2 Using Games and Game Elements beyond Entertainment

When using games and game elements beyond entertainment, Deterding et al. [De11] propose three basic usage types: Gamification, serious games and games with a purpose. These usage types are described in the following.

Gamification can be defined as using game design elements in non-game contexts [De11]. Common game elements include points, badges, leaderboards and avatars [ZC11]. The aim of gamification often lies in motivating a specific behavior of users by implementing different (mostly social) reward structures [Ka12]. Considering this intention, gamification is defined in this paper as using game elements in non-game contexts in order to motivate a specific user behavior.

In contrast to using just game elements, serious games employ full-fledged games [De11]. They are often defined as games that are not limited to the purpose of entertainment [Ab87]. Serious games originate from the military, they are hence mostly concerned with acquiring new skills and teaching players educational content [Sm10]. In opposition to "educational games" (or "edugames"), this educational content can hardly be separated from the game mechanics, which is why learning takes place while playing

the game [Ch10]. Debriefing is also an important activity that fosters reflection on the content when using serious games [GAD02]. In order to accommodate the different purposes of serious games, they are defined in this paper as games that aim at entertaining players as well as improving their skills or knowledge.

As a last usage type of games and game elements beyond entertainment, "games with a purpose" are defined as games that solve real-world problems just by being played [De11]. They are being used for example by biologists to predict protein structures with the collective intelligence of players [Co10].

3 Method

For conducting a scientifically sound literature review about the theoretical foundation of using games and game elements for learning and motivation, this study employs the review setup suggested by Fettke [Fe06]. This categorization can be used to clarify the characteristics of a review study and is based on several recommendations from literature [Fe06]. According to this framework, this study presents a review in natural language that focuses on theory, takes a neutral perspective and highlights central aspects on the basis of selective literature (cf. Tab. 1).

Characteristic		Category			
1. Type		natural language		mathematical-statistical	
2. Focus		research result	research method	theory	experience
3. Aim	Mention	not mentioned		mentioned	
	Content	integration	criticism		central aspects
4. Perspective		neutral		position	
5. Literature	Selection	not explained		explained	
	Scope	key works	representative	selective	exhaustive
6. Structure		chronologically	thematically		methodically
7. Target Audience		common public	practitioners	researchers in general	specialized researchers
8. Future Research		not mentioned		mentioned	

Tab. 1: Characterization of this review based on Fettke [Fe06]

Games are gaining increasing relevance in the Information Systems (IS) domain as so-called "interactive hedonic systems" [LB10]. This review therefore starts with examin-

ing publications from this domain. Thus, a manual search of relevant research articles in the AIS Senior Scholars' Basket of Journals is conducted. Additional sources for the manual search are the journal "Decision Support Systems" (DSS) and a special issue on gamification in "Creativity and Innovation Management" (CAIM). Conference proceedings from the "International Conference on Information Systems" (ICIS) and the annual meeting on informatics in the "Lecture Notes in Informatics" (LNI) are also included in the search. The AIS Senior Scholars' Basket of Journals is considered since it comprises a widely accepted set of top journals in the field of IS research. The journals DSS and CAIM are selected because of their relevance to business and information systems engineering. Last, the conference proceedings from the ICIS and LNI are considered since they provide current publications from manifold research communities, such as human computer interaction. The investigation period covers the years 2009 to 2014. Every title as well as (in case of relevant terms) every heading is searched for formulations which indicate that games or game elements are being used beyond entertainment. The result of this manual search consists of 42 relevant publications.

In a next step, journals for a structured literature search are identified by looking up the references of the relevant publications mentioned above for journals that are specialized on the usage of games and game elements. This identification revealed the journals "Simulation & Gaming" (S&G) and "Games and Culture" (G&C). These journals are therefore being used for a structured keyword search considering all publications until 2014. The search terms employed are the usage types of games and game elements beyond entertainment presented by Deterding et al. [De11]. Since the term "Serious Gaming" is often used synonymously with "Serious Games", both search terms are used for serious games. The search terms are depicted in Tab. 2.

Usage type	Search term(s)
Serious Games	("Serious" AND "Games") OR ("Serious" AND "Gaming")
Gamification	"Gamification"
Games with a purpose	"Games" AND "purpose"

Tab. 2: Search terms employed in the keyword search in the journals S&G and G&C

Keyword search in the journals S&G and G&C revealed another 25 relevant publications. Together with the publications already identified by manual search, the literature sample consists of 67 publications. Since Ping et al. [PGT10] and Goh and Ping [GP14] report about the same study, only the more recent publication is considered for the literature sample. Hence, 66 publications remain in the literature sample for this review.

4 Theoretical Foundation of Using Games and Game Elements for Learning and Motivation

In the literature sample investigated, we can distinguish between theory-based publications and non-theory-based publications. A publication is hereby called theory-based, when it explains how games and game elements facilitate learning and motivation by referring to theories. Only 34 publications can be called theory-based according to this definition. They name 28 different theories, of which 6 are mentioned more than twice. Since this indicates their relevance in the field, these theories are selected and presented in Tab. 3.

Theory	Focus	Number of Mentions
Flow Theory of Motivation	Motivation	17
Self-determination Theory	Motivation	10
Experiential Learning Theory	Learning	6
Goal-setting Theory	Motivation	4
Bloom's Taxonomy	Learning	3
Constructivist Learning Theory	Learning	3

Tab. 3: Theories mentioned more than once in the literature sample

In order to characterize the theoretical foundation of using games and game elements for learning and motivation, these theories as well as the extent to which they are employed in the respective publications are described in the following.

4.1 Flow Theory of Motivation

Flow theory of motivation is the most cited theory in the literature sample. It describes a so-called "flow state", in which people forget about their surroundings and lose their sense of time [Cs91]. The state of flow is characterized by intense concentration, merging of action and awareness, loss of reflective self-consciousness, a sense that one can control one's actions, distortion of temporal experience and experience of an activity as intrinsically rewarding [NC02]. When being in flow, an individual operates at full capacity which means they even neglect hunger, fatigue or discomfort in order to continue pursuing an activity [NC02]. However, in order to experience the flow state, the challenge of the activity has to be in balance with the skills of the individual: Too much challenge causes anxiety, whereas too little challenge leads to boredom [NC02]. Flow has been described as a part of the gameplay experience in the literature sample [Al12; Be12; DBS13; Ha13; Ka12; KH13; LLS13; Mu11; Na12; OL13; Ok13; PR14; RUO14; SW13; WS14; WR12; WSR11]. Bedwell et al. [Be12] link various game attributes to learning outcomes. The game attribute "conflict/challenge" can thus lead to a flow state, if the degree of challenge automatically adapts to the skill level of the player. Kankan-

halli et al. [Ka12] reference Chen [Ch07] who states that flow is important for a game experience. The challenge of a game therefore has to match the player's skills [Ka12]. Koops and Hoevenaar [KH13] note that serious games are likely to trigger a flow state. However, they argue that flow might even distract the player from learning, since deeper reflection on the content of the game does not take place while being in flow [KH13]. Liu et al. [LLS13] agree that challenge has to match the player's skills. However, since their publication focuses on competition in games, they define the challenge of a game as the opposing player's skills. Hence, they conclude that both players' skill levels have to match in order to enter the flow state [LLS13]. Mueller et al. [Mu11] examine the use of virtual worlds as knowledge management platforms. In their study, they found that users of virtual worlds reported a flow-like state [Mu11]. They hence propose that because of the game-like characteristics of a virtual world, a flow state is achieved which in turn leads to important knowledge-related activities [Mu11]. Nadolski et al. [Na12] investigate architectures for multiuser learning scenarios and declare flow as the optimal learning state. They conclude that it is important for these architectures to ensure a flow state e.g. by logging player data in order to inform design [Na12]. Oksanen [Ok13] refers to flow as one of the seven core game experiences during gameplay and agrees that challenge has to match the player's skills in order to enter the flow state. The remaining publications also mention flow as a part of the gameplay experience and important for player motivation [Al12; DBS13; Ha13; OL13; PR14; RUO14; SW13; WS14; WR12; WSR11]. In summary, flow can be seen as a core experience of gameplay and is achieved by the challenge of a game corresponding to the player's individual skills.

4.2 Self-determination Theory

The main aspects of self-determination theory are motivation and personality, thus it can be called a motivation theory [RD00]. Different psychological needs are a central construct of this theory [RD00]. Every human thus has the need for competence, relatedness and autonomy. Fulfilling these needs leads to motivation, whereas neglecting them results in discouragement [RD00]. In the literature sample, publications mention self-determination theory to describe motivational effects of gamified applications [Ka12; LHC12; LLS13; MK14; Sc15; SW13; Te13; WS14; WR12; WSR11]. However, self-determination theory is not a crucial element in most of these publications; it is rather mentioned among others in their literature overviews. Only Kankanhalli et al. [Ka12] and Liu et al. [LLS13] link the psychological needs of self-determination theory with digital video games by referencing Ryan et al. [RRP06]. Following this argumentation, autonomy is achieved in games by letting players choose sequences of actions [RRP06]. Perceived competence is enhanced by tasks within the game that provide optimal challenges, and a feeling of relatedness can be achieved for example in multiplayer games, where players interact with each other [RRP06]. Taking a look at competitive elements, Liu et al. [LLS13] note that competition can have both positive and negative impacts on the enjoyment of the gameful experience: External incentives might for example undermine the feeling of autonomy, since the player is pushed into a certain direction. However, they also claim that competition can satisfy the player's need for competence

[LLS13]. Mutter and Kundisch [MK14] agree that external rewards like badges can lower the player's perceived autonomy. While this may lead to an increase in the quantity of player contribution, the contribution quality might suffer [MK14]. To sum it up, self-determination theory can be linked to video games in general, and also describes how intrinsic motivation in gamified applications can be achieved [Sc15; SW13; WS14; WR12; WSR11].

4.3 Experiential Learning Theory

Experiential learning theory underlines the influence of experience on learning success [Ko84] and can hence be called a learning theory. A central construct of this theory is the so-called learning cycle which is composed of concrete experience, reflective observation, abstract conceptualization and active experimentation [Ko84]. These stages are attached to corresponding activities (i.e., feeling, watching, thinking, doing) which result in different learning styles (i.e., diverging, assimilating, converging, and accommodating) [Ko84]. Diverging refers to individuals who prefer feeling and watching (e.g., looking at concrete situations from several different viewpoints), assimilating embraces watching and thinking (e.g., thinking through logical explanations for observed phenomena), converging covers doing and thinking (e.g., applying theoretical knowledge to practical applications), and accommodating incorporates doing and feeling (e.g., trying things out rather than thinking them through) [Ko84]. Taking into account these learning styles might lead to more effective learning [Ko84]. The publications in the literature sample argue that experiential learning is supported by interactivity in games [Al12; Be10; KH13; Le13; ML11; Na12]. In contrast to other approaches in the sample, Koops and Hoevenaar [KH13] directly incorporate elements from experiential learning theory as a part of their "Serious Gaming Lemniscate Model" which consists of a learning cycle and a gaming cycle. Hereby, they provide a link between flow theory of motivation and experiential learning theory: While the gaming cycle corresponds to an experience similar to flow, the learning cycle is consistent with the learning cycle in experiential learning theory. The authors argue that by manipulating a game's difficulty, a transition between the gaming cycle (i.e., the flow state) and the learning cycle takes place [KH13]. Hence, their model provides a first link between learning and motivation theories in the literature sample [KH13]. Monk and Lycett [ML11] describe a modified (which means strongly simplified) version of experiential learning theory by using a learning cycle that consists only of act, reflect and understand. Alklind Taylor et al. [Al12] further simplify experiential learning theory by only stating that practical experience has to precede theoretical discussion of educational content. Nadolski et al. [Na12] cite Kebritchi and Hirumi [KH08] who link the pedagogical foundations of learning games with experiential learning. Ben-Zvi [Be10] and Legner et al. [Le13] call games a form of experiential learning without further justification. We summarize that experiential learning can take place in serious games, as long as they provide possibilities to go through the stages of the learning cycle.

4.4 Goal-setting Theory

Goal-setting theory describes how goals influence motivation and task performance of individuals. Locke and Latham [LL02] draw on 35 years of empirical research on goal-setting theory, pointing out goal mechanisms and moderators of goal effects. The four goal mechanisms described consist of a directive function (goals direct attention toward goal-related activities and away from goal-irrelevant activities), an energizing function (high goals lead to greater effort than low goals), goals affecting persistence (hard goals prolong effort) and goals affecting action indirectly (goals lead to the arousal, discovery, and/or use of task-relevant knowledge and strategies). Especially the last mechanism hints at goals also leading to learning outcomes. One of the most important moderators of goal effects is goal commitment [LL02]. High goal commitment leads to a strong goal-performance relationship. The more difficult the goal, the more commitment is needed. Goal commitment is supported by the perceived importance of the goal. This perceived importance can be raised e.g. by individuals making a public commitment to the goal or letting them choose their own goals. Self-efficacy is also important for goal commitment, especially when it comes to difficult goals. It can be raised by providing success experiences, finding role-models to identify with, and persuasive communication that the individual can reach the goal (e.g., by providing solution strategies). The remaining moderators of goal effects are feedback (revealing progress in relation to the goals), task complexity (high complexity of the goal requires the ability to discover appropriate task strategies), personal goals as mediators of external incentives (i.e., taking into account personal goals and self-efficacy of a person when assigning goals), and satisfaction (achieving goals leads to satisfaction). In the literature sample, goal-setting theory is used to describe why players want to achieve certain accomplishments in gamified applications [Ha13; MK14; Op14; SW13]. Mutter and Kundisch [MK14] investigate a gamified Q&A-community and argue that goal-setting theory applies to badges in gamified applications, since badges can resemble a valuable goal to players, mostly because of their function as status symbols. Oppong-Tawiah et al. [Op14] propose that using specific, difficult and obtainable goals has a strong effect in persuasive gamified applications that foster pro-environmental behavior. Haas et al. [Ha13] and Scheiner and Witt [SW13] do not refer directly to goal-setting theory, but to self-efficacy, which is part of goal-setting theory. These publications therefore provide a theory-based explanation how goals in gamified applications lead to player motivation and performance.

4.5 Bloom's Taxonomy

Bloom's taxonomy describes several consecutive steps of the cognitive process that consist of remembering, understanding, applying, analyzing, evaluating and creating [AK01]. The publications in the literature sample use Bloom's taxonomy to describe the learning outcomes of serious games [Be10; Le13; ML11]. Ben-Zvi [Be10] proposes Bloom's taxonomy as an assessment framework for the outcomes of experiential learning. Monk and Lycett [ML11] suggest using serious games as a way of testing knowledge in higher education by aiming at the higher levels of Bloom's taxonomy

(e.g., analyzing and evaluating). Last, Legner et al. [Le13] link the learning outcomes of business simulation games with Bloom's taxonomy. In summary, Bloom's taxonomy is used to define and to assess the learning outcomes in serious games.

4.6 Constructivist Learning Theory

Constructivist learning theory is rather a philosophical view of comprehension and knowledge in general [SD95]. A central construct of this theory is the view that knowledge does not exist on its own but is constructed in each individual's mind [SD95]. Several so-called constructivist learning methods stem from this theory [SD95]. The publications in the literature sample mention these constructivist learning methods in combination with serious games [Ch10; Na08; Th06]. Charsky [Ch10] claims that more and more experts call for constructivist learning methods. Some of these learning methods can thus be fulfilled by serious games. A scientific evaluation of this claim, however, has still to be executed [Ch10]. Thomas [Th06] mentions constructivist learning as one of the concepts on which their so-called "pervasive learning" is based. Nadolski et al. [Na08] report a high demand for constructivist learning methods. They justify the link between serious games and constructivist learning methods only by educational experts' opinions that serious games can meet this demand. The literature sample does therefore not contain any publication that rigorously links serious games with constructivist learning methods or that references such a publication.

4.7 Summary and Discussion

The literature review disclosed 6 relevant theories used to explain how games and game elements facilitate learning and motivation. It also showed to which extent they are incorporated in the literature sample, reaching from "mentioned in the literature overview" to "substantial part of the publication". For most of the theories, a sufficient link to games and game elements is provided. As mentioned above, this is not the case for constructivist learning theory. One possible explanation for this is the broad scope of this theory, being rather a philosophical view of understanding and knowledge in general. Although basically being a promising possibility for future theorizing, this theory is taken out of consideration in this paper. The remaining learning theories (i.e., experiential learning theory and Bloom's taxonomy) show the different stages of the learning cycle (i.e., concrete experience, reflective observation, abstract conceptualization, and active experimentation) and provide a framework for categorizing and assessing the desired learning outcomes (i.e., remembering, understanding, applying, analyzing, evaluating, and creating). The motivation theories (i.e., flow theory of motivation, self-determination theory, and goal-setting theory) show that player motivation depends on the challenge of a gameful experience corresponding to the player's skills, the player perceiving competence, autonomy, and relatedness, as well as the player trying to reach several goals.

5 Theory-driven Design Guidelines for Game-based Learning and Motivation

In the following, 10 design guidelines for game-based learning and motivation are derived from the theoretical insights provided by the literature review. They result from checking every theory presented in section 4 for ways to enable learning and motivation. A short listing of the suggested design guidelines is presented in Tab. 4.

Design Guideline	Theoretical Foundation
Balance challenge and skill	Flow Theory
Enable perceived competence, autonomy, and relatedness	Self-determination Theory
Employ all stages of the learning cycle	Exp. Learning Theory
Consider different learning styles	Exp. Learning Theory
Set specific, difficult, and obtainable goals	Goal-setting Theory
Enable perceived goal importance	Goal-setting Theory
Enable goal-related self-efficacy	Goal-setting Theory
Constantly show progress in relation to goals	Goal-setting Theory
Remind of accomplished goals	Goal-setting Theory
Categorize and assess specific learning outcomes	Bloom's Taxonomy

Tab. 4: Theory-driven design guidelines for game-based learning and motivation

The first design guideline is called "balance challenge and skill" and corresponds to flow theory of motivation. Thus, the challenge of a game has to match the player's skills in order to enter the flow state (see section 4) [NC02]. This can for example be achieved by automatically adapting the level of difficulty in a game [Be12] or by matching players with equal skill levels in a competitive setting [LLS13]. Players might also be given the possibility to choose a difficulty level by themselves, which may in addition lead to higher goal commitment (cf. goal-setting theory). For example, a player that chooses a "hard" difficulty setting might feel more obliged to beat the game, than when given no choice. Being able to choose an "easy" difficulty setting may on the other hand help players with low self-efficacy to gain confidence about being able to beat the game.

As described in section 4, games can foster self-determination and intrinsic motivation by providing the feeling of competence, autonomy, and relatedness [RRP06]. It is therefore important for the design of such gameful experiences to promote these feelings, which is expressed by the design guideline "enable perceived competence, autonomy, and relatedness". Autonomy is achieved in games by letting players choose sequences of actions [RRP06]. Perceived competence is enhanced by tasks within the game that provide optimal challenges and a feeling of relatedness can be achieved for example in multiplayer games, where interactions between players can take place [RRP06].

Experiential learning theory supports two design guidelines. The design guideline "employ all stages of the learning cycle" suggests that a gameful experience should encompass every learning activity (i.e., concrete experience, reflective observation, abstract conceptualization, and active experimentation) in order for players to go through the entire learning cycle (cf. section 4). This can be achieved in games by incorporating different gameplay mechanics (e.g., providing a notepad) and addressing these activities in debriefing. Taking into account different learning styles, for example by offering separate game modes for assimilating (thinking-oriented) or accommodating (action-oriented) learning styles, is addressed by the design guideline "consider different learning styles".

Goal-setting theory supports several design guidelines. The first one is to "set specific, difficult, and obtainable goals", since it has been shown that setting a specific and difficult goal leads to higher performance than simply urging players to do their best [LL02]. These specific goals can for example be badges or quests in gameful experiences. As perceived goal importance leads to higher goal commitment, an additional guideline is to "enable perceived goal importance". This can be done by letting players make public commitments to their goals (e.g., by showing their goals to other players) or letting them choose their own goals. Another guideline is to "enable goal-related self-efficacy", since this also leads to higher goal commitment. Self-efficacy can be raised by providing success experiences (e.g., easier goals for beginners), presenting role-models (e.g., players close in a leaderboard or a coaching system with experienced players), and persuading players that they are able to reach a goal (e.g., by providing hints). According to goal-setting theory, it is also important to "constantly show progress in relation to goals", e.g. by providing progress bars or quest logs. Since accomplishing goals leads to satisfaction, another idea is to "remind of accomplished goals" to raise player satisfaction.

To help clearly point out the desired learning outcomes of gameful experiences, Bloom's Taxonomy can be utilized. Before the development of an application/game, the desired learning outcomes may be mapped to the different steps of the cognitive process. This is due to the fact that learning outcomes like "remembering" might need different game mechanics than for example developing the ability to "evaluate". When assessing the learning outcomes after a gaming session, this categorization can be used again to see if the players actually acquired the respective capabilities [Be10]. This design guideline is referred to as "categorize and assess specific learning outcomes".

In summary, designers of game-based learning and motivation should pay attention to the underlying mechanisms that lead to learning and motivation in order to develop successful applications. It is important to note, however, that these design guidelines are not meant to be mandatory, hence not every design guideline has to be fully executed in every application for game-based learning and motivation.

6 Conclusion and Future Research

This study provides a basic understanding of how games and game elements facilitate learning and motivation. It also presents 10 theory-driven design guidelines for game-based learning and motivation. In contrast to existing approaches, it examines both motivation and learning theories. Since this study is a first step towards identifying the theoretical foundation of using games and game elements for learning and motivation, a limitation of these results is the restriction of the search space. It is in the nature of this research method, that one single literature review can impossibly cover all relevant publications that exist on the topic. However, this does not affect the usefulness of this paper, since it only means that future works can add to the theoretical foundation identified by this review. Future research may also focus on developing a specific theory for serious games or gamification e.g. by combining the presented learning and motivation theories. In addition, the suggested design guidelines can be empirically evaluated, e.g. by examining if existing games are violating these guidelines and how this affects the intended outcomes. After this evaluation, the presented design guidelines can be used in future research to develop scientifically sound serious games and gamified applications.

Bibliography

[Ab87] Abt, C. C.: *Serious Games.* University Press of America, 1987.

[AK01] Anderson, L. W.; Krathwohl, D. R.: *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman, New York, 2001.

[Al12] Alklind Taylor, A.-S.; Backlund, P.; Niklasson, L.: *The Coaching Cycle: A Coaching-by-Gaming Approach in Serious Games*. In: Simulation & Gaming 43/5, p. 648–672, 2012.

[ALB06] von Ahn, L.; Liu, R.; Blum, M.: *Peekaboom: a Game for Locating Objects in Images*. In (Grinter, R. et al. Eds.): Proceedings of the SIGCHI conference on Human Factors in computing systems. Montréal, p. 55–64, 2006.

[Be10] Ben-Zvi, T.: *The efficacy of business simulation games in creating Decision Support Systems: An experimental investigation*. In: Decision Support Systems 49/1, p. 61–69, 2010.

[Be12] Bedwell, W. L. et al.: *Toward a Taxonomy Linking Game Attributes to Learning: An Empirical Study*. In: Simulation & Gaming 43/6, p. 729–760, 2012.

[Bo10] Borrajo, F. et al.: *SIMBA: A simulator for business education and research*. In: Decision Support Systems 48/3, p. 498–506, 2010.

[Bo15] Bogost, I.: *Why Gamification Is Bullshit*. In (Walz, S. P.; Deterding, S. Eds.): The Gameful World - Approaches, Issues, Applications. MIT Press, Cambridge, MA, p. 65–80, 2015.

[Ch07] Chen, J.: *Flow in games (and everything else)*. In: Communications of the ACM 50/4, p. 31, 2007.

[Ch10] Charsky, D.: *From Edutainment to Serious Games: A Change in the Use of Game Characteristics*. In: Games and Culture 5/2, p. 177–198, 2010.

[Co10] Cooper, S. et al.: *Predicting protein structures with a multiplayer online game*. In: Nature 466/7307, p. 756–760, 2010.

[Cs91] Csikszentmihalyi, M.: *Flow: The psychology of optimal experience.* HarperPerennial New York, 1991.

[DBS13] Döpker, A.; Brockmann, T.; Stieglitz, S.: *Use Cases for Gamification in Virtual Museums*. In (Horbach, M. Eds.): Informatik 2013. Ges. für Informatik, Bonn, p. 2308–2320, 2013.

[De11] Deterding, S. et al.: *From game design elements to gamefulness: defining "gamification"*. In (Lugmayr, A. et al. Eds.): MindTrek '11: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, p. 9–15, 2011.

[Fe06] Fettke, P.: *State-of-the-Art des State-of-the-Art*. In: Wirtschaftsinformatik 48/4, p. 257–266, 2006.

[GAD02] Garris, R.; Ahlers, R.; Driskell, J. E.: *Games, Motivation, and Learning: A Research and Practice Model*. In: Simulation & Gaming 33/4, p. 441–467, 2002.

[GP14] Goh, K.-Y.; Ping, J. W.: *Engaging Consumers with Advergames: An Experimental Evaluation of Interactivity, Fit and Expectancy*. In: Journal of the Association for Information Systems 15/7, p. 388–421, 2014.

[Ha13] Haas, P. et al.: *Der Einfluss von Gamification auf die empfundene Selbstwirksamkeit von Teilnehmern von Online-Ideengenerierungswettbewerben über die Zeit*. In (Horbach, M. Eds.): Informatik 2013. Ges. für Informatik, Bonn, p. 2321–2335, 2013.

[Ka12] Kankanhalli, A. et al.: *Gamification: A New Paradigm for Online User Engagement*. In (Huang, M.-H.; Piccoli, G.; Sambamurthy, V. Eds.): Proceedings of the 33rd International Conference on Information Systems (ICIS). Shanghai, p. 1–10, 2012.

[KH08] Kebritchi, M.; Hirumi, A.: *Examining the pedagogical foundations of modern educational computer games*. In: Computers & Education 51/4, p. 1729–1743, 2008.

[KH13] Koops, M.; Hoevenaar, M.: *Conceptual Change During a Serious Game: Using a Lemniscate Model to Compare Strategies in a Physics Game*. In: Simulation & Gaming 44/4, p. 544–561, 2013.

[Ko84] Kolb, D. A.: *Experiential learning: Experience as the source of learning and development.* Prentice-Hall Englewood Cliffs, NJ, 1984.

[LB10] Lin, C.-P.; Bhattacherjee, A.: *Extending technology usage models to interactive hedonic technologies: a theoretical model and empirical test*. In: Information Systems Journal 20/2, p. 163–181, 2010.

[Le13] Legner, C. et al.: *Designing Capstone Courses in Management Education: Knowledge Activation and Integration Using an ERP-based Simulation Game*. In (Baskerville, R.; Chau, M. Eds.): Proceedings of the 34th International Conference on Information Systems (ICIS). Milan, p. 1–19, 2013.

[LHC12] Li, Z.; Huang, K.-W.; Cavusoglu, H.: *Quantifying the Impact of Badges on User Engagement in Online Q&A Communities*. In (Huang, M.-H.; Piccoli, G.; Sambamurthy, V. Eds.): Proceedings of the 33rd International Conference on Information Systems (ICIS). Shanghai, p. 1–10, 2012.

[LL02] Locke, E. A.; Latham, G. P.: *Building a Practically Useful Theory of Goal Setting and Task Motivation*. In: American Psychologist 57/9, p. 705–717, 2002.

[LLS13] Liu, D.; Li, X.; Santhanam, R.: *Digital Games and Beyond: What Happens When Players Compete*. In: MIS Quarterly 37/1, p. 111–124, 2013.

[MK14] Mutter, T.; Kundisch, D.: *Behavioral Mechanisms Prompted by Badges: The Goal-Gradient Hypothesis*. In (Myers, M.; Straub, D. Eds.): Proceedings of the 35th International Conference on Information Systems (ICIS). Auckland, p. 1–17, 2014.

[ML11] Monk, E.; Lycett, M.: *Using a Computer Business Simulation to Measure Effectiveness of Enterprise Resource Planning Education on Business Process Comprehension*. In (Beath, C.; Myers, M.; Wei, K.-K. Eds.): Proceedings of the 32nd International Conference on Information Systems (ICIS). Shanghai, p. 1–10, 2011.

[Mu11] Mueller, J. et al.: *Virtual worlds as knowledge management platform - a practice-perspective*. In: Information Systems Journal 21/6, p. 479–501, 2011.

[Na08] Nadolski, R. J. et al.: *EMERGO: A methodology and toolkit for developing serious games in higher education*. In: Simulation & Gaming 39/3, p. 338–352, 2008.

[Na12] Nadolski, R. J. et al.: *Architectures for Developing Multiuser, Immersive Learning Scenarios*. In: Simulation & Gaming 43/6, p. 825–852, 2012.

[NC02] Nakamura, J.; Csikszentmihalyi, M.: *The concept of flow*. In (Snyder, C. R.; Lopez, S. J. Eds.): Handbook of positive psychology. Oxford University Press, New York, NY, USA, p. 89–105, 2002.

[Ok13] Oksanen, K.: *Subjective Experience and Sociability in a Collaborative Serious Game*. In: Simulation & Gaming 44/6, p. 767–793, 2013.

[OL13] Oberdörfer, S.; Latoschik, M. E.: *Develop your strengths by gaming: Towards an inventory of gamificationable skills*. In (Horbach, M. Eds.): Informatik 2013. Ges. für Informatik, Bonn, p. 2346–2357, 2013.

[Op14] Oppong-Tawiah, D. et al.: *Encouraging Sustainable Energy Use in the Office with Persuasive Mobile Information Systems*. In (Myers, M.; Straub, D. Eds.): Proceedings of the 35th International Conference on Information Systems (ICIS). Auckland, p. 1–11, 2014.

[PGT10] Ping, J.; Goh, K.; Teo, H.: *Engaging Consumers with Advergames: An Experimental Evaluation of Interactivity, Relevancy and Expectancy*. In (Lacity, M.; March, S.; Niederman, F. Eds.): Proceedings of the 31st International Conference on Information Systems (ICIS). St Louis, 2010.

[PR14] Plennert, S.; Robra-Bissantz, S.: *Ein Serious Game als neue qualitative Erhebungsmethode für die Customer Journey beim Automobilkauf*. In (Plödereder, E. et al. Eds.): Informatik 2014. Köllen, Bonn, p. 383–398, 2014.

[RD00] Ryan, R. M.; Deci, E. L.: *Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being*. In: American Psychologist 55/1, p. 68, 2000.

[RRP06] Ryan, R. M.; Rigby, C. S.; Przybylski, A.: *The Motivational Pull of Video Games: A Self-Determination Theory Approach*. In: Motivation and Emotion 30/4, p. 344–360, 2006.

[RUO14] Romero, M.; Usart, M.; Ott, M.: *Can Serious Games Contribute to Developing and Sustaining 21st Century Skills?* In: Games and Culture 10/2, p. 1–30, 2014.

[Sc15] Scheiner, C. W.: *The Motivational Fabric of Gamified Idea Competitions: The Evaluation of Game Mechanics from a Longitudinal Perspective*. In: Creativity and Innovation Management 24/2, p. 341–352, 2015.

[SD95] Savery, J. R.; Duffy, T. M.: *Problem Based Learning: An instructional model and its constructivist framework*. In: Educational Technology 35/5, p. 31–38, 1995.

[Sm10] Smith, R.: *The Long History of Gaming in Military Training*. In: Simulation & Gaming 41/1, p. 6–19, 2010.

[SW13] Scheiner, C. W.; Witt, M.: *The Backbone of Gamification - a Theoretical Consideration of Play and Game Mechanics*. In (Horbach, M. Eds.): Informatik 2013. Ges. für Informatik, Bonn, p. 2372–2386, 2013.

[Te13] Teh, N. et al.: *Can Work Be Fun? Improving Task Motivation and Help-Seeking Through Game Mechanics*. In (Baskerville, R.; Chau, M. Eds.): Proceedings of the 34th International Conference on Information Systems (ICIS). Milan, 2013.

[Th06] Thomas, S.: *Pervasive learning games: Explorations of hybrid educational gamescapes*. In: Simulation & Gaming 37/1, p. 41–55, 2006.

[WR12] Witt, M.; Robra-Bissantz, S.: *Sparking Motivation and Creativity with "Online Ideation Games"*. In (Goltz, U. et al. Eds.): Informatik 2012. Ges. für Informatik, Bonn, p. 1006–1023, 2012.

[WS14] Wiegand, T.; Stieglitz, S.: *Serious Fun-Effects of Gamification on Knowledge Exchange in Enterprises*. In (Plödereder, E. et al. Eds.): Informatik 2014. Köllen, Bonn, p. 321–332, 2014.

[WSR11] Witt, M.; Scheiner, C. W.; Robra-Bissantz, S.: *Gamification of Online Idea Competitions: Insights from an Explorative Case*. In (Heiß, H.-U. et al. Eds.): Informatik 2011. Ges. für Informatik, Bonn, p. 1–15, 2011.

[WW02] Webster, J.; Watson, R. T.: *Analyzing the past to prepare for the future: Writing a literature review*. In: MIS Quarterly 26/2, p. 3, 2002.

[ZC11] Zichermann, G.; Cunningham, C.: *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps.* O'Reilly Media, Inc., Sebastopol, CA, 2011.

[ZHR12] van der Zee, D.-J.; Holkenborg, B.; Robinson, S.: *Conceptual modeling for simulation-based serious gaming*. In: Decision Support Systems 54/1, p. 33–45, 2012.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

GamEducation in einer virtuellen 3D-Umgebung mit Googles Virtual-Reality-Brille Cardboard

Linda Eckardt[1], Jan-Paul Huttner[2], Susanne Robra-Bissantz[3]

Abstract: Spielelemente wurden in den vergangenen Jahren bereits als Motivationsanreiz innerhalb von Lehrveranstaltungen an Hochschulen eingesetzt. Oftmals werden dabei den Studierenden fiktive Charaktere (z.B. Vampire) als Avatare zur Verfügung gestellt. Einige Studierende können sich mit diesen Figuren nicht identifizieren. GamEducation, wobei Spielelemente in der Lehre eingesetzt werden und Studierende innerhalb einer realitätsnahen Umgebung als realistische Charaktere agieren, wirkt dem entgegen. Auch das Lernen über Virtual-Reality-Brillen bietet Potential zur Verbesserung des Lernverhaltens. Dennoch gibt es bislang kaum Anwendungen auf diesem Gebiet. Der vorliegende Beitrag beschreibt daher ein Konzept, bei dem Studierende mit einer Virtual-Reality-Brille in einer realitätsnahen und virtuellen Lernumgebung in einem gamifizierten Kontext, interaktiv und mit direktem Feedback in konkreten Anwendungsszenarien lernen.

Keywords: Gamification, Education, GamEducation, Head-Mounted Display, Virtual Reality

1 Einleitung und Motivation

Spielen nimmt beim Erwerb von Fähigkeiten eines Menschen im Verlauf des Heranwachsens, insbesondere im Kleinkindalter, eine zentrale Rolle ein. Allerdings verändert sich das Verständnis des Lernens beim Erwachsenwerden von einem Lernen auf spielerische Art und Weise zu einer Auffassung des Lernens als ernsthafte Angelegenheit [Sh11]. Die Anwendung von Gamification kann dem entgegenwirken. Unter Gamification wird die Integration von Spielelementen (z.B. Spielpunkte, Ranglisten, Geschichten) in einen spielfremden Kontext verstanden [De11]. Demnach ist Gamification nicht mit einem Spiel gleichzusetzen. Vielmehr wird die durch Spielelemente hervorgerufene Motivation versucht in anderen Bereichen nutzbar zu machen. Beispielsweise können Punkte oder Abzeichen als Anreize für die Erledigung von alltäglichen Aufgaben vergeben werden. Durch diese zusätzlichen Motivationsanreize werden Nutzer länger an eine Anwendung gebunden und setzen sich folglich auch intensiver mit ihr auseinander [ZC11]. Existierende Ansätze in der Hochschullehre nutzen bereits Gamification, integrieren darüber hinaus den

[1] Technische Universität Braunschweig, Institut für Wirtschaftsinformatik, Lehrstuhl Informationsmanagement, Mühlenpfordtstraße 23 (5. OG), 38106 Braunschweig, linda.eckardt@tu-bs.de
[2] Technische Universität Braunschweig, Institut für Wirtschaftsinformatik, Lehrstuhl Informationsmanagement, Mühlenpfordtstraße 23 (5. OG), 38106 Braunschweig, j-p.huttner@tu-bs.de
[3] Technische Universität Braunschweig, Institut für Wirtschaftsinformatik, Lehrstuhl Informationsmanagement, Mühlenpfordtstraße 23 (5. OG), 38106 Braunschweig, s.robra-bissantz@tu-bs.de

individuellen Wissensstand und zeigen den persönlichen Lernfortschritt auf [KS13].

Obwohl in vielen Veröffentlichungen darauf verwiesen wird, dass auch virtuelle Umgebungen über Virtual-Reality-Brillen (VR-Brillen) ein Wachstumsfeld mit hohem Potential zur Veränderung und Verbesserung der Lernerfahrungen sind, gibt es hierzu bisher kaum konkrete Anwendungen in der Lehre [Pa95, DE09]. Hier wurde beispielsweise eine Anwendung zur Vermittlung von biologischen Konzepten als virtuelle Realität implementiert, um Lernenden die Struktur und Funktion von Zellen näher zu bringen [GG94]. In einem weiteren Beispiel lernen Schüler das Designen von Wurzeln, Stielen und Blättern von Pflanzen, um in verschiedenen Umgebungen zu wachsen [MM04]. Allerdings sind gängige Anwendungen in diesem Bereich auf die Lernerfahrung einzelner Personen ausgelegt, d.h. der Lernende wird in einem abgegrenzten Raum mit Sensoren zur Erfassung von physischen Bewegungen und einem Head-Mounted-Display (HMD) ausgestattet [Pa06]. Folglich findet also oftmals ein isoliertes Lernen statt. Zudem sind die bisherigen HMDs, wie die Oculus Rift mit einem Preis von ca. vierhundert Euro, zu teuer für einen breiten Einsatz. Die einfach zu transportierende und einzusetzende VR-Brille Cardboard von Google, mit einem Preis von ca. zehn Euro, gestattet erstmals einen relativ kostengünstigen und breiten Einsatz in der Lehre, außerhalb von abgegrenzten Räumen und damit einhergehend mit der Möglichkeit der Kooperation mit weiteren Lernenden.

In diesem Beitrag wird ein Konzept vorgestellt, bei dem Studierende in einer realitätsnahen und virtuellen Lernumgebung in einem gamifizierten Kontext, interaktiv und mit direktem Feedback in konkreten Anwendungsszenarien die wesentlichen (informations-)technischen und wirtschaftswissenschaftlichen Grundlagen der Wirtschaftsinformatik kennen lernen.

2 GamEducation

Die Anwendung von Gamification in der Lehre ist nicht neu. Beispielsweise wird die „Legende von Zyren“, ein programmiertes Fantasy-Rollenspiel, dazu genutzt, Studierenden der Universität Düsseldorf auf spielerische Art und Weise Wissensrepräsentationen beizubringen [KS13]. Die Studierenden wählen zwischen Elfen, Orks, Menschen und Goblins als Repräsentation ihres Charakters und lösen sowohl einzeln, als auch als Gruppe in sogenannten Gilden Aufgaben auf einer begleitenden Plattform oder in einer Präsenzveranstaltung [KS13]. Ein weiteres Beispiel für Gamification in der Lehre stellt „Creatures of the Night“, konzipiert für die Vermittlung von Kenntnissen in der Mathematik an der Hochschule Heidelberg, dar [KR14]. Als Vampire oder Werwölfe sollen die Studierenden im Verlauf des Semesters einerseits durch die Teilnahme an Clantreffen (Übungen) und Ratsversammlungen (Vorlesungen), andererseits durch das Absolvieren von Aufgaben auf einer begleitenden Plattform, Punkte und Abzeichen für entsprechende Leistungen sammeln [KR14].

Die Verwendung von fiktiven Geschichten in Verbindung mit fiktiven Figuren als

Repräsentation der eigenen Person führt bei einem Teil der Studierenden dazu, sich nicht ernst genommen zu fühlen [KR14]. Aus diesem Grund wird in dem in diesem Beitrag beschriebenen Ansatz neben der Nutzung von weiteren Spielelementen der Fokus auf eine begleitende Geschichte gelegt, mit der sich Studierende aufgrund ihres Alltags identifizieren können. Zudem repräsentieren die Studierenden selbst ihre virtuelle Identität. Dieses Vorgehen beschreibt GamEducation. Der Begriff GamEducation ist eine Kombination aus den Wörtern Gamification und Education und bezeichnet demnach den Einsatz von Spielelementen in der Lehre, wobei die Lernenden innerhalb von realitätsnahen Geschichten als realistische Charaktere agieren.

An der technischen Universität Braunschweig wird GamEducation in der Lehrveranstaltung „Kooperationen im E-Business" bereits seit dem Wintersemester 2012/13 eingesetzt. Die Vermittlung von Vorlesungsinhalten wird durch eine Praxisaufgabe (Hintergrundgeschichte) begleitet, innerhalb derer Studierende als Mitarbeiter eines Unternehmens im gegenseitigen Wettbewerb an einem Konzept zur Erschließung neuer Märkte arbeiten. Die Praxisaufgabe ermöglicht den stetigen Transfer und die Anwendung des Vorlesungswissens. Darüber hinaus sammeln die Studierenden in mehreren Challenges sowohl als Gruppe, als auch einzeln Punkte. Auf einem lehrveranstaltungsbegleitenden Blog diskutieren die Studierenden Inhalte und sehen die aktuelle Rangliste als Feedback auf die eigene Leistung. Lehrevaluationen haben gezeigt, dass das Konzept die Mehrheit der Studierenden dazu motiviert, sich in deutlich höherem Maße zu beteiligen und die Anwendung des Vorlesungswissens dabei hilft, die Inhalte besser zu verstehen.

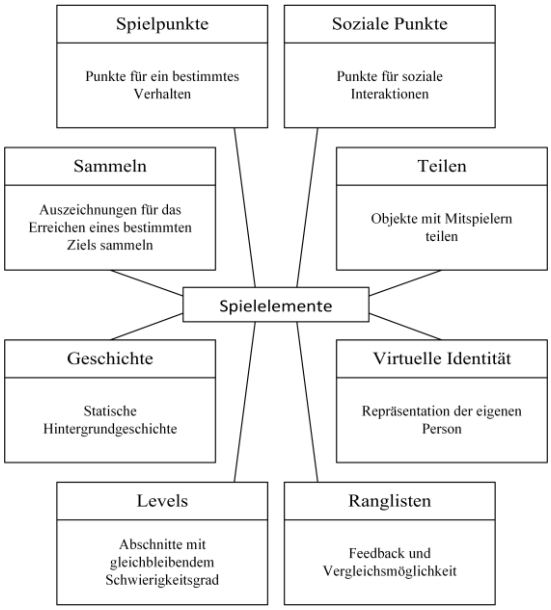


Abb. 1: Spielelemente in GamEducation mit der VR-Brille Cardboard

GamEducation ist nicht auf bestimmte Spielelemente festgelegt und variiert daher in Abhängigkeit des gewählten Anwendungsfalls. In Abbildung 1 sind die Spielelemente in Anlehnung an Kim (2009) die innerhalb des in diesem Beitrag beschriebenen Konzepts zum Lernen mit der VR-Brille Cardboard eingesetzt werden, dargestellt [Ki09].

Geschichte

Geschichten treten in statischer oder dynamischer Form auf [Ki09]. Beide Varianten helfen den Spielern beim Erzeugen eines mentalen Modells und fördern in einigen Fällen die Motivation, da Spieler die nächsten Geschehnisse im Spielverlauf wissen wollen [KBM13]. Statische Geschichten, wie in dem Konzept, sind vordefiniert, passiv und stellen oftmals als Hintergrundgeschichte einen narrativen Rahmen bereit, in dem Spieler handeln können [MW00]. Dynamische Geschichten hingegen sind ungeplant und interaktiv, d.h. Spieler haben die Möglichkeit aktiv den Spielverlauf zu beeinflussen [Cr04].

Spielpunkte

Punkte sind, unabhängig davon, ob nur für einen selbst oder auch für andere Teilnehmer sichtbar, ein wesentliches Spielelement [ZC11]. Die Berücksichtigung einer Punktevergabe beim Designen einer Gamification-Anwendung ist notwendig und daher existiert eine Auswahl von verschiedenen Punktesystemen [ZC11]. Im Rahmen dieses Beitrags stehen Spielpunkte und soziale Punkte im Fokus. Teilnehmer erhalten für ein bestimmtes Verhalten automatisch Spielpunkte vom System [HA09]. Diese Punkte dienen als unmittelbares Feedback und da alle Spieler die gleiche Punktzahl für eine bestimmte Aktivität erhalten, führt das Punktesystem zu einer Form des Wettbewerbs und Vergleichs untereinander [Wi13].

Soziale Punkte

Im Gegensatz zu den Spielpunkten, werden soziale Punkte nicht vom System vergeben, sondern von anderen Mitspielern als positive Beurteilungen [Ki09]. Absolute und relative Bewertungen werden dabei unterschieden [FÜ10]. Ersteres erlaubt Spielern die Vergabe einer bestimmten Punktzahl und Letzteres ermöglicht über einen paarweisen Vergleich die Zuordnung von Punkten [FÜ10]. Innerhalb des in diesem Beitrag beschriebenen Konzepts erhalten Studierende eine feste Anzahl von Punkten von ihren Mitspielern.

Levels

Die Anhäufung von Punkten ermöglicht das Entwerfen von Levels und Ranglisten. Mehrere Arten von Levels können unterschieden werden. Beim Level- oder Missionsbasierten Design, wie in dem in diesen Beitrag beschriebenen Konzept, muss in jedem Level eine kleine Menge an Zielen erfüllt werden, bevor der Spieler in das nächste Level wechseln kann [Ka12]. Durch dieses Design ist der Spielraum für den Spieler zu bewältigen und wirkt nicht einschüchternd [Ka12].

Ranglisten

Ranglisten sind ein weiteres Spielelement und zeigen den Spielern ihren Rangplatz. Das Ranking ermöglicht den Spielern einen Vergleich untereinander und dient als Feedback für die eigenen Erfolgschancen [RR09].

Virtuelle Identität

Die virtuelle Identität ist die Repräsentation der eigenen Person innerhalb der Spielumgebung, mit der Mitspieler interagieren können [Ca05]. Hat der Nutzer die Möglichkeit, seine virtuelle Identität anzupassen, beispielsweise hinsichtlich des Alters, der Haarfarbe, der Kleidung oder des Geschlechts, wird die virtuelle Identität Avatar genannt [Ji09]. Spieler neigen dazu Avatare zu wählen, die optische Ähnlichkeit mit der eigenen Person haben, sich allerdings charakterlich von einem selbst unterscheiden [HLW5, BSK07]. Dadurch wird dem Spieler die Möglichkeit gegeben eine eigene Identität innerhalb der Spielumgebung zu kreieren, die frei von Schwächen aus dem echten Leben ist, dafür aber Anonymität berücksichtigt [BSK07, TR10].

Sammeln

Das Sammeln von z.B. Auszeichnungen ist ein weiteres Spielelement [Ki09]. Auszeichnungen haben die Funktion, den Spielern den stetigen Fortschritt innerhalb des Systems zu zeigen und markieren das Erreichen bestimmter Ziele [ZC11]. Darüber hinaus fördert das Sammeln eine Form des Wettbewerbs, da Spieler danach streben ihre Sammlung an Auszeichnungen zu vervollständigen und diese Mitspielern zu zeigen, die sie verstehen oder sogar bewundern [ZC11, DK89, Fo91].

Teilen

Teilen als Element von Austauschprozessen stellt ein weiteres Spielelement dar [Ki09]. In dem in diesem Beitrag beschriebenen Konzept ist das Teilen eine Form der Kooperation, beruht auf Gegenseitigkeit und kann das Ansehen eines Spielers positiv beeinflussen [SLH06].

3 GamEducation mit der Virtual-Reality-Brille Cardboard

Nachfolgend wird das Konzept zum gamifizierten Lernen mit der VR-Brille Cardboard vorgestellt. Lernen in virtuellen Umgebungen mithilfe einer VR-Brille kann dem Konstruktivismus zugeordnet werden [Ma01]. Die Erkenntnistheorie des Konstruktivismus besagt, dass Lernen ein selbstgesteuerter Prozess ist, bei dem sich die Lernenden durch eine Interpretation der persönlichen Wahrnehmungserfahrungen eine individuelle Repräsentation der Welt erschaffen [Jo91]. In virtuellen Umgebungen haben die Studierenden die Möglichkeit zu interagieren und einzutauchen. Dadurch ist ein Lernen auf Basis von eigenen Erfahrungen und in den Situationen mit den entsprechenden Lerninhalten gegeben. Daneben ist eine direkte Interaktion mit Inhalten

für das Aufbauen eines eigenen Verständnisses beim Lernen wesentlich. Die Präsentation von vorstrukturierten Inhalten durch Dozenten oder Texte, wie es in gängigen Lehrveranstaltungen der Fall ist, ist im Vergleich weniger für das Lernen geeignet [Se80]. Durch die direkte Interaktion innerhalb der virtuellen Welt wird eine individuelle Repräsentation des Wissens konstruiert. Dadurch werden die zuvor in der Vorlesung vermittelten theoretischen Grundlagen (deklaratives Wissen) verständlicher und erfahrbarer (prozedurales Wissen).

3.1 Virtual-Reality-Brille Cardboard

Google hat die VR-Brille Cardboard im Sommer 2014 auf der I/O Entwicklerkonferenz vorgestellt [CPS14]. Es handelt sich dabei um einen Bausatz für eine VR-Brille. Der Bausatz umfasst u.a. zurechtgeschnittene Pappteile, Linsen für die richtige Optik und einen Magnetschalter. Zusammengebaut entsteht ein Gehäuse für das eigene Smartphone, um dieses als VR-Display zu nutzen. In Abbildung 2 ist der Aufbau in Anlehnung an Jannsen (2015) dargestellt [Ja15].

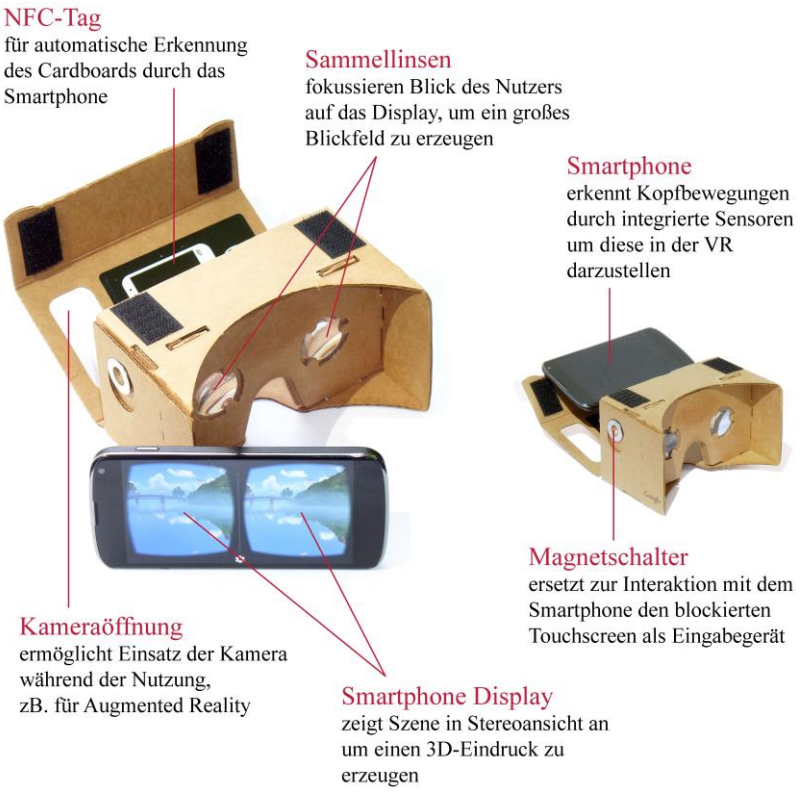


Abb. 2: Aufbau der VR-Brille Cardboard

Die Benutzung von VR-Brillen kann Symptome, wie Schwindel oder Kopfschmerzen auslösen (Simulatorkrankheit) [La00]. Als Auslöser können neben der visuellen Simulation allerdings auch andere Faktoren ausschlaggebend sein [KF92]. Cardboard ermöglicht dem Nutzer nachdem installieren einer entsprechenden App auf dem Smartphone das Eintauchen in eine dreidimensionale und virtuelle Welt [Go15]. In Kombination mit den Sensoren des Smartphones kann die Kopfbewegung erfasst und folglich genutzt werden, um sich in einer virtuellen Welt umzuschauen [Go15]. Darüber hinaus ermöglicht der Magnetschalter ein Klicken in der App, d.h. dieser kann als Eingabemöglichkeit verwendet werden, um in der virtuellen Umgebung zu interagieren und somit zu lernen.

3.2 GamEducation mit Google's Cardboard

Die Studierenden können demnach durch die Cardboards eine virtuelle Welt erkunden und darin lernen. Ein Haus mit einem Raum pro Vorlesungsthema repräsentiert die virtuelle Umgebung für den einzelnen Studierenden. In jedem Raum können Spielpunkte gesammelt werden. Dazu erkunden die Studierenden in der Egoperspektive die einzelnen Räume. Die Möglichkeit aus eigener Erfahrung zu lernen soll dazu beitragen das Verhalten der Studierenden auch außerhalb der virtuellen Umgebung positiv zu beeinflussen und Gelerntes auf andere Situationen zu übertragen [Ka12].



Abb. 3: Beispiel für einen virtuellen Raum (mit Blender erstellt)

Die Abbildung 3 zeigt einen prototypischen Aufbau für so einen virtuellen Raum, in dem sich die Studierenden durch die Cardboards umsehen und darin interagieren können. Der eBook Reader auf dem Tisch kann in diesem ersten Beispiel das Themenfeld der digitalen Medienprodukte, wie z.B. elektronische Bücher oder digitale Musik, repräsentieren. Das bedeutet, dass der Studierende den eBook Reader auswählen kann und virtuell anhand des Lesens eines digitalen Buches die Eigenschaften digitaler Produkte tatsächlich und realitätsnah anhand verschiedener Handlungsalternativen erfährt: beispielsweise könnte der Studierende das Buch kopieren oder mit einem anderen Studierenden teilen. Dies würde dann die Eigenschaften leichte Übertrag- und Duplizierbarkeit repräsentieren. Der Studierende könnte aber auch Textstellen virtuell markieren und dadurch die Eigenschaft der leichten Veränderbarkeit erfahren.

Demnach erkunden die Studierenden die virtuelle Welt individuell. Dennoch bilden sie untereinander je Raum zufällige Gruppen. Für jedes Mitglied einer Gruppe wird es pro Raum unterschiedliche Objekte wie z.B. den eBook Reader zur Repräsentation eines bestimmten Themengebiets geben. Dafür, dass die Studierenden innerhalb ihrer virtuellen Umgebung mit einem Objekt interagieren, wird es Punkte geben. Darüber hinaus gibt es für den Austausch von Objekten innerhalb der zufälligen Gruppe soziale Punkte. Dadurch ist gewährleistet, dass sich die Studierenden untereinander unterstützen und auf eine gewisse Weise kooperieren. Ohne diesen Austausch untereinander, ist es nicht möglich alle in dem Raum zugehörigen Objekte und folglich alle erfahrbaren Beispiele für ein bestimmtes Themengebiet zu erhalten.

Räume werden in der Woche freigeschaltet, in der die Inhalte in der Vorlesung vermittelt werden. Falls in dem vorherigen Raum bereits eine Mindestanzahl an Punkten erreicht wurde, wird der nächste Raum bereits vor der Vorlesung freigeschaltet. Dieses Vorgehen soll die Studierenden dazu motivieren Lernzeiten individuell und überall auszunutzen, um die Inhalte zu verstehen. Darüber hinaus wird der individuelle Lernfortschritt durch eine Erweiterung der virtuellen Umgebung erkennbar. Das bedeutet, bei einem Lernfortschritt werden immer mehr neue Räume und Orte in der Welt sichtbar. Entsprechend wird dem Studierenden sein Erfolgsfeedback unmittelbar in einer einprägsamen räumlichen Dimension präsentiert. Außerdem ermöglicht das Freischalten des nächsten Raums bzw. Levels zu Vorlesungsbeginn, dass die Dozenten im Anschluss an ein bestimmtes Themengebiet innerhalb einer Vorlesung dazu auffordern können die Cardboards aufzusetzen und das Gelernte nochmals auf erfahrbare Art und Weise zu verinnerlichen.

Durch die Objekte innerhalb eines Raumes zur Repräsentation eines Vorlesungsthemas wird der Mechanismus des Gedächtnispalastes (Loci-Methode) im Lernprozess eingesetzt und die Studierenden können die virtuelle Welt, nach der Übungsphase, in Gedanken durchschreiten, um die benötigten Informationen zu memorieren. Die Loci-Methode dient dem Zweck, sich an viele Informationen und Zusammenhänge erinnern zu können, wobei die zu erinnernden Informationen mit Bildern und Orten verknüpft werden [Sp84]. Entsprechend werden diese Bilder an bestimmten Positionen mental verankert und nicht mehr bewegt. Um sich später an die abgelegten Bilder bzw.

Informationen zu erinnern, läuft der Lernende seinen Gedächtnispalast in Gedanken Raum für Raum ab [Sp84]. Untersuchungen haben bereits gezeigt, dass die Präsentation eines virtuellen Raumes an einem Computer in seiner Eignung und Effektivität einer in Gedanken selbst entworfenen oder bereits vertrauten Umgebung in nichts nachsteht, um für die Loci-Methode eingesetzt zu werden [LE12].

Da es für jede Interaktion mit Objekten und in Abhängigkeit des Austauschs von Objekten mit anderen Studierenden Punkte gibt, wird es in jedem Raum für den Studierenden im oberen Bereich des Blickfelds sichtbar eine Auszeichnung, mit den Abstufungen Bronze, Silber und Gold, geben. Die Bronze-Auszeichnung erscheint, wenn mit allen im Raum bereits vorhandenen Objekten interagiert wurde. Die Silber-Auszeichnung erscheint bei einem Objektaustausch und einer anschließenden Interaktion mit diesem Objekt. Die Gold-Auszeichnung gibt es, wenn alle möglichen weiteren Objekte in diesem Raum gesammelt und ausprobiert wurden. Neben diesen Auszeichnungen gibt es eine vereinfachte Rangliste, die neben der eigenen Platzierung die Rangplätze über und unter einem anzeigt, um einen direkten Vergleich zu ermöglichen und die Motivation zu fördern, da der Studierende erkennt, dass der nächste Rangplatz aufgrund des geringen Punkteunterschieds zu erreichen ist.

4 Schlussbemerkungen

Insgesamt bietet das in diesem Beitrag beschriebene Lehrvorhaben eine durch das Cardboard zu erkundende virtuelle Umgebung, in der das Gelernte/die Wissensbasis virtuell einprägsam durch die Räume und darin befindlichen Objekte präsentiert wird und der Weg zum Wissenserwerb gamifiziert erfolgt.

Nachdem Gamification in der Hochschullehre bereits weit verbreitet ist und aus Untersuchungen hervorging, dass sowohl eine Förderung der Motivation, als auch eine intensivere Auseinandersetzung mit den Lerninhalten erfolgt, bietet dieses Konzept durch die Verbindung mit der VR-Brille Cardboard eine neue Anwendungsmöglichkeit für Gamification in der Lehre. Dabei können die Studierenden entsprechend ihrer Bedürfnisse Lernzeiten individuell und überall ausnutzen, um Vorlesungsinhalte zu vertiefen oder zu wiederholen.

Im nächsten Schritt erfolgt die Umsetzung des geplanten Konzepts. Dazu werden die Objekte der virtuellen Umgebung modelliert und in eine Spiel-Engine importiert, um Navigation und Interaktionen innerhalb der Umgebung zu ermöglichen.

Im Anschluss an die Realisierung soll innerhalb von empirischen Untersuchungen gezeigt werden, ob Lernen in virtuellen 3D-Umgebungen mit einer VR-Brille und unter Berücksichtigung von Gamification einen Mehrwert für das Lernerlebnis schafft, das Lernverhalten positiv beeinflusst und die Lernmotivation erhöht.

Literaturverzeichnis

[BSK07] Bessière, K.; Seay, F.; Kiesler, S.: The ideal elf. Identity exploration in world of warcraft. CyberPsychology and Behavior 4/07, S. 530-535, 2007.

[Ca05] Castranovy, E.: Synthetic Worlds. The Business and Culture of Online Games. The University of Chicago Press, Chicago, 2005.

[Cr04] Crawford, C.: Chris Crawford on Interactive Storytelling. New Riders, Berkeley, 2004.

[CPS14] Coz, D.; Plagemann, C. ;Smus, B. , https://www.google.com/events/io/io14videos, Stand: 17.06.2015.

[DK89] Danet, B.; Katriel, T.: Not wo alike: Play and aesthetics in collecting. Play and Culture 3/89, S. 253-277, 1989.

[De09] Dede, C., Immersive Interfaces for Engagement and Learning, Science 323/5910, S. 66-69, 2009.

[De11] Deterding, S. et.al.: Gamification. Toward a Definition. Paper presented at the CHI 2011 Workshop Gamification, VanCouver, Canada.

[Fo91] Formanek, R.: Why they collect. Collectors reveal their motivations. In (Rudmin, F.W. Hrsg.): To Have Possessions. A Handbook on Ownership and Property. Select Press, Corte Madera, 1991.

[FÜ10] Füller, J. et.al.: Evaluation games. How to make the crowd your jury. In (Fähnrich, K. P.; Franyzyk, B. Hrsg.): Lecture Notes Informatics Proceedings, S. 955-960, 2009.

[GG94] Gray E.; Greschler, D.: Is virtual reality a good teaching tool. Boston Computer Museum, 1994.

[Go15] Google, https://developers.google.com/cardboard/overview, Stand: 17.06.2015.

[HA09] Hacker, S.; Ahn, L.: Matchin. Elicting user preferences with online game. In Proc. 27th Int. Conf. on human factors in computing systems, Boston, Messachussets, S. 1207-1216, 2009.

[HLW05] Hsu, S. H.; Lee F. L.; Wu, M.C.: Designing action games for appealing to buyers. CyberPsychology and Behavior 6/05, S. 585-591, 2005.

[Ja15] Jannsen, J., http://www.heise.de/ct/ausgabe/2015-7-Das-Smartphone-wird-zur-Virtual-Reality-Brille-2562140.html, Stand: 17.06.2015.

[Ji09] Jin, S. A.: Avatars Mirroring the Actual Self versus Projecting the Ideal Self. The Effects of Self-Priming on Interactivity and Immersion in an Exergame Wii Fit. CyberPsychology & Behavior 6/09, S. 761-765, 2009.

[Jo91] Jonassen, D. H.: Objectivism versus constructivism. Do we need a new philosophical paradign. Educational technology research and development 3/91, S. 5-14, 1991.

[Ka12] Kapp, K. M.: The Gamification of learning and instruction. Game-based methods and strategies for training and education. John Wiley & Sons, San Francisco, 2012.

[KBM13] Kapp, K. M.; Blair, L.; Mesch, R.: The gamification of learning and instruction fieldbook. Ideas into practice. John Wiley & Sons, San Francisco, 2013.

[KF92] Kennedy, R. S.; Fowlkes, J. E.: Simulator sickness is polygenic and polysymptomatic: Implications for research. The International Journal of Aviation Psychology 1/92, S. 23-38, 1992

[Ki09] Kim, A. J.: Putting the Fun in Functional. Applying Game Mechanics to Social Media. Presented at Startup2Startup, Palo Alto, California, 2009.

[KS13] Knautz, K.; Sabousta, S.: Aufbruch nach Zyren. Game-based Learning in der Hochschullehre. Universitätsbibliothek Hildesheim, 2013.

[KR14] Kruse, V. et.al.: Creatures of the Night: Konzeption und Evaluation einer Gamification-Plattform im Rahmen einer Mathematikvorlesung. 2014.

[La00] LaViola Jr, J. J.: A discussion of cybersickness in virtual environments. ACM SIGCHI Bulletin 1/00, S. 47-56, 2000

[LE12] Legge, E. L. G. et.al.: Building a memory palace in minutes. Equivalent memory performance using virtual versus conventional environments with the method of loci. Jornal of Acta Psychologica 3/12, S. 380-390, 2012.

[Ma01] Mantovani, F.: VR Learning. Potential and Challenges for the Use of 3D Environments in Education and Training. In (Riva, G.; Galimberti, C. Hrsg.): Towards CyberPsychology. Mind, Cognitions and Society in the Internet Age, IOS Press, Amsterdam, 2001.

[MM04] Moreno, R.; Mayer, R. E.: Personalized Messages That Promote Science Learning in Virtual Environments. Journal of Educational Psychology 96/01, S. 165-173, 2004.

[MW00] Mallon, B.; Webb, B.: Structure, causality, visibility and interaction. Propositions for evaluating engagement in narrative multimedia. International Journal of Human-Computer-Studies 02/00, S. 269-287, 2000.

[Pa95] Pantelidis, V. S.: Reasons to Use Virtual Reality in Education. VR in the Schools 1/95, 1995.

[Pa06] Patel, K.et.al.: The effects of fully immersive virtual reality on the learning of physical tasks. In Proc. of the 9th Annual Workshop on Presence, Ohio, USA, 2006.

[RR09] Reeves, B.; Read, J. L.: Total Engagement. Using Games and Virtual Worlds to Change the Way People Work and Business Complete. McGraw-Hill Professional, New York, 2009.

[Se80] Seymour, P.: Mindstorms. Children, computers, and powerful ideas. Basic Books, 1980.

[Sh11] Sheldon, L.: The multiplayer classroom: Designing coursework as a game. Cengage Learning, 2011.

[Sp84] Spence, J. D.: The Memory Palace of Matteo Ricci, Penguin Books, 1984.

[SLH06] Sun, C. T.; Lin, H.; Ho C. H.: Sharing tips with strangers. Exploiting gift culture in computing gaming. CyberPsychology and Behavior 5/06, S. 560-570, 2006.

[TR10] Trepte, S.; Reinecke, L.: Avatar creation and video game enjoyment. Journal of Media Psychology 4/10, S. 171-184, 2010.

[Wi13] Witt, M.: Application of Game Mechanics to Innovation Management: Theoretical Foundations and Empirical Studies. Ph.D thesis, University of Braunschweig, 2013.

[ZC11] Zichermann, G.; Cunningham, C.: Gamification by Design. Implementing Game Mechanics in Web and Mobile Apps. O'Reilly, Köln, 2011.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Teilhabe an der allgegenwärtigen Kommunikation

Teilhabe an der allgegenwärtigen Kommunikation

Arbeitskreis „Inklusion in Social Media“

Computer werden allgegenwärtig. Sie versprechen unsere Sinne zu erweitern, unser Gedächtnis zu unterstützen, uns von Arbeit zu entlasten und unsere Sicherheit zu erhöhen. Dinge werden mit einander interagieren und uns von Alltäglichem befreien. Wie können diese technischen Entwicklungen dazu beitragen, alle Menschen mit und ohne spezielle Bedürfnisse am beruflichen und gesellschaftlichen Leben teilhaben zulassen. In unserem Workshop wollen wir Möglichkeiten kennenlernen und zukünftigen Entwicklungsoptionen im Hinblick auf die Teilhabe aller am Internet der Dinge miteinander diskutieren. Ebenso wird die Barrierefreiheit von Tangible User Interfaces und Natural User Interfaces Thema unseres Workshops sein. Dazu möchten wir Forschende, Lehrende und Entwickelnde aus Theorie und Praxis einladen.

Programmkomitee

Christian Beck, Zentrum für angewandte assistierende Technologien Wien

Prof. Dr. Christian Bühler, TU Dortmund

Prof. Dr. Ricardo Büttner, FOM Hochschule für Oekonomie & Management München

Wiebke Köhlmann, Institut für Informatik, Universität Potsdam

Dr. Florian Kramer, SignGes, RWTH Aachen

Henning Lübbecke, FH Bund

Dr. Steffen Puhl, Hochschulrechenzentrum, Justus-Liebig-Universität Gießen

Dr. René Reiners, Fraunhofer Institute for Applied Information Technology, St. Augustin

Prof. Dr. Rainer Stiefelhagen, Karlsruher Institut für Technologie

Prof. Dr. Gerhard Weber, TU Dresden

Prof. Dr. Gottfried Zimmermann, Hochschule der Medien Stuttgart

Francis Zinke, Institut für Informatik, Universität Potsdam

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Förderung der medialen Teilhabe von Menschen mit erworbenen Hirnschädigungen durch assistive Technologien

Susanne Dirks[1] Nadezda Anton[2] Christian Bühler[3]

Abstract: Im Rahmen des hier vorgestellten Projektes soll ein digitales Assistenzsystem entwickelt werden, mit dessen Hilfe Menschen mit einer erworbenen Hirnschädigung weitgehend barrierefrei und eigenständig am digitalen Leben teilhaben können. Mit Hilfe einer Nutzerbedarfsanalyse wurde in ersten Projektphase untersucht, ob und in welcher Weise Menschen mit erworbenen Hirnschädigungen die neuen Medien als Mittel zur Teilhabe am öffentlichen Leben nutzen und mit welchen Barrieren sie dabei konfrontiert werden. Die Auswertung der Ergebnisse zeigte, dass die Nutzung der neuen Medien nach dem Eintritt der Hirnschädigung insgesamt zurückgegangen ist, obwohl nach wie vor ein großer Wunsch besteht, diese im Sinne einer umfassenden Teilhabemöglichkeit zu nutzen. Auf Basis der Ergebnisse der Nutzerbedarfsanalyse soll in der folgenden Projektphase ein Prototyp für einen integrierten Arbeitsplatz entworfen werden, an dem die betroffene Nutzergruppe weitgehend eigenständig unter Nutzung verschiedener assistiver Hilfsmittel die Kommunikations- und Informationsangebote des Internets und digitale Therapie- und Unterhaltungssoftware nutzen kann.

Keywords: Assistive Technologien, Barrierefreie Softwareentwicklung, Teilhabe, AT für Menschen mit Behinderungen, erworbene Hirnschädigung

1 Einleitung

Das Internet und die neuen Medien werden zunehmend nicht nur für Recherchen und Büroarbeiten genutzt, sondern gehören in vielfältiger Weise zum alltäglichen Leben dazu. In den Wirtschaftsrechnungen des Statistischen Bundesamtes zur Nutzung von Informations- und Kommunikationstechnologien für das Jahr 2014 [DES15] wird angegeben, dass 95% aller Haushalte mit 3 und mehr Personen einen Computer, Laptop oder Tablet PC besitzen und 84% aller Haushalte über einen leistungsfähigen Internetanschluss verfügen. Je jünger die in einem Haushalt lebenden Personen sind und je höher ihr Bildungsstand ist, desto größer ist die Verbreitung der Internetnutzung (bis zu 100% bei Schülern und Studierenden). Durch die Möglichkeit der Internetnutzung mit mobilen Endgeräten und die Verfügbarkeit günstiger Flatrate-Tarife werden auch die Alltagskommunikation, der Konsum und die Unterhaltung zunehmend ins Internet verlegt.

Obwohl viele Internetangebote mittlerweile auf Grund gesetzlicher Bestimmungen weitgehend barrierefrei für Menschen mit sensorischen oder motorischen Beeinträchtigungen

[1] TU Dortmund, TIP-Cluster / Rehabilitationstechnologie, Emil-Figge-Str. 50, 44227 Dortmund, susanne.dirks@tu-dortmund.de
[2] TU Dortmund, Forschungsstelle Teilhabe / Rehabilitationstechnologie, Emil-Figge-Str. 50, 44227 Dortmund, nadezda.anton@tu-dortmund.de
[3] TU Dortmund, TIP-Cluster / Rehabilitationstechnologie, Emil-Figge-Str. 50, 44227 Dortmund, christian.buehler@tu-dortmund.de

sind, sind Menschen mit kognitiven Beeinträchtigungen von diesen Entwicklungen immer noch weitgehend ausgeschlossen. Nur wenige Internetangebote präsentieren Informationen in Leichter Sprache oder ermöglichen den Nutzern, ablenkende Seiteninhalte auszuschalten oder die Menge der auf einer Seite angebotenen Informationen einzuschränken. Menschen mit einer angeborenen oder erworbenen Hirnschädigung haben vielfältige Probleme bei der Nutzung neuer Medien [GD15]. Neben erheblichen motorischen Einschränkungen sind oft auch die kommunikativen und kognitiven Fähigkeiten sowie Aufmerksamkeits- und Gedächtnisfunktionen temporär oder dauerhaft beeinträchtigt.

Die 2008 in Kraft getretene UN-Behindertenrechtskonvention sieht für alle Menschen eine uneingeschränkte Teilhabe am öffentlichen und politischen Leben und das Recht auf eine selbstbestimmte Lebensführung vor. Menschen mit einer erworbenen Hirnschädigung sind auf Grund der starken Heterogenität ihrer Störungen und der vorher meist unabhängigen Lebensführung mit besonders großen Veränderungen konfrontiert. Untersuchungen zur Verwendung moderner Technologien im Alltag und im beruflichen Umfeld durch ältere Menschen und Menschen mit erworbenen Hirnschädigungen, haben gezeigt, dass der Wunsch Alltagstechnologie zu nutzen nach wie vor stark ausgeprägt ist, die Fähigkeit zur Nutzung aber durch die vorliegenden körperlichen und kognitiven Beeinträchtigungen sowie Unzulänglichkeiten in der Nutzerfreundlichkeit der Endgeräte stark eingeschränkt ist [KPLL13].

Das hier beschriebene Projekt ist konzeptuell in den Rahmen des bio-psycho-sozialen Modells funktionaler Gesundheit eingebettet. Die WHO bietet mit der International Classification of Functioning, Disability and Health (ICF) eine fachübergreifende einheitliche und standardisierte Sprache zur Beschreibung des funktionalen Gesundheitszustandes, der Behinderung, der sozialen Beeinträchtigung und der relevanten Umgebungsfaktoren eines Menschen. Mit der ICF können die bio-psycho-sozialen Aspekte von Komponenten der Gesundheit unter Berücksichtigung der Kontextfaktoren systematisch erfasst werden.

Abbildung 1 zeigt das der ICF zugrunde liegende Verständnis der Wechselwirkungen zwischen den verschiedenen Komponenten. Die Fähigkeit eines Menschen Funktionen in einer spezifischen Domäne auszuführen ist als eine komplexe Beziehung zwischen dem spezifischen Gesundheitsproblem und verschiedenen Kontextfaktoren (Umweltfaktoren und personenbezogene Faktoren) zu verstehen. Es besteht eine dynamische Wechselwirkung zwischen diesen Größen: Interventionen bezüglich einer Größe können eine oder mehrere der anderen Größen verändern. Diese Wechselwirkungen sind spezifisch, stehen aber nicht immer in einem vorhersagbaren Eins-zu-Eins-Zusammenhang. So ist es z.B. möglich, dass bei einigen Personen durch eine eingeschränkte Teilhabe und die daraus resultierenden eingeschränkten Aktivitäten Gesundheitsbeeinträchtigungen entstehen oder verstärkt werden können, während andere Personen durch einen eingeschränkte Teilhabe, beispielsweise im Medienbereich, keine Aktivitätseinschränkung empfinden und demzufolge auch nicht mit deren Auswirkungen auf den Gesundheitszustand konfrontiert sind.

Mit dem Ziel, die Nutzungsbedürfnisse und die aktuelle Mediennutzung bei Menschen mit erworbenen Hirnschädigungen zu erfassen und ihre mediale Teilhabe zu fördern, wurde im Rahmen der 2008 begonnen Kooperation „2be in action“ zwischen Bethel.regional und der Fakultät Rehabilitationswissenschaften der TU Dortmund im August 2013 das



Abb. 1: Komponenten der Gesundheit nach der ICF (siehe [De05])

Projekt „Förderung der medialen Teilhabe von Menschen mit Behinderungen durch assistive Technologien“ ins Leben gerufen. Im ersten Projektabschnitt wurde eine Nutzerbedarfsanalyse zu den Möglichkeiten und Risiken der medialen Teilhabe in Einrichtungen der Bethel.regional - Regionen Dortmund und Ruhr sowie der gGmbH „In der Gemeinde leben“ Düsseldorf durchgeführt. Das Ergebnis der Studie bietet eine Grundlage zur Konzeption einer bedarfsorientierten Förderung der medialen Teilhabe bei Menschen mit Unterstützungsbedürfnissen nach einer erworbenen Hirnschädigung.

Im Folgenden werden die Ergebnisse der Nutzerbedarfsanalyse genauer vorgestellt und im Hinblick auf die Entwicklung eines prototypischen Arbeitsplatzes zur eigenständigen Nutzung von geeigneten Endgeräten und Softwareprodukten diskutiert.

2 Nutzerbedarfsanalyse

Im Rahmen der Nutzerbedarfsanalyse sollten die aktuelle Kommunikationssituation und die Kommunikationsbedürfnisse von Menschen mit erworbenen Hirnschädigungen untersucht werden. Basierend auf den Fähigkeiten und den zur Verfügung stehenden Hilfsmitteln der definierten Zielgruppe wurde zunächst ein Instrument erstellt, mit dessen Hilfe die aktuelle Kommunikationssituation und die Mediennutzung sowie das Verbesserungspotential der medialen Teilhabe erhoben werden konnten.

2.1 Methodik

Menschen mit erworbenen Hirnschädigungen bilden hinsichtlich ihrer Funktionseinschränkungen, ihren daraus resultierenden Unterstützungsbedürfnissen und auch hinsichtlich der vorhandenen Ressourcen eine sehr heterogene Gruppe. Die Annahmen über die Spannbreite der Beeinträchtigungen und der vorhandenen Ressourcen wurden deshalb von der Koordinatorengruppe als nicht abschließend einschätzbar bewertet und aus diesem Grund

wurde ein klientenzentriert-partizipatives Design für die Feldstudie gewählt, d.h. die Probanden und ihre Betreuer wurden eng in die Entwicklung des Fragebogens eingebunden.

2.2 Untersuchungsinstrument

Im Projekt wurden in Zusammenarbeit mit potentiellen Probanden und deren Betreuer nach dem Vorbild des Fragebogens zur Nutzung von Alltagstechnologie [MKNa13] strukturierte Fragebögen zur Mediennutzung und zur Lebenszufriedenheit bezogen auf die mediale Teilhabe und die persönliche Kommunikationssituation entworfen. Die Bezugsbetreuer der Probanden wurden in der Anwendung der Fragebögen geschult und sollten ihre Klienten dabei unterstützen die Fragen selbständig oder mit geringer Hilfe zu beantworten. Bei der Konzeption des Fragebogens wurden darüber hinaus die Richtlinien der Leichten Sprache [Bu14] berücksichtigt und die Kriterien zur Lese- und Interpretierbarkeit von Texten von Huenerfauth et al. [HFE09] umgesetzt.

Zusätzlich zur Papierumfrage wurde eine Online-Version des Fragebogens mit Hilfe der Umfragesoftware Limesurvey erstellt. Abbildung 2 zeigt einen Ausschnitt der Online-Umfrage.



Abb. 2: Onlineumfrage - Beispielscreen

2.3 Probanden

Der für die Nutzerbedarfsanalyse ausgewählte Personenkreis umfasste Klienten der Stiftung Bethel.regional in den Regionen Dortmund und Ruhr und der gGmbH „In der Gemeinde leben“ Düsseldorf. Zu Studienbeginn wurden die Verantwortlichen der Einrichtungen, die Klienten der Zielgruppe und deren rechtliche Betreuer über die Inhalte und Ziele des Projektes informiert. Die Bereitschaft zur Teilnahme an der Studie wurde mit einer Einverständniserklärung sowie einer Erklärung zur Schweigepflicht und zur Datensicherheit in vereinfachter Sprache dokumentiert.

2.4 Datenerhebung

Als Methode für die Datenerhebung wurde eine Befragung, bestehend aus verschiedenen Frage- und Antworttypen eingesetzt. Es wurden Informations- und Funktionsfragen eingesetzt, die und er Regel durch Alternativ- und Multiple-Choice-Auswahlen beantwortet werden konnten. Zur Erhebung der Nutzungswünsche wurden offene Fragen eingesetzt, die die Probanden offen beantworten konnten. Da die Probanden freiwillig an der Studie teilnahmen und entweder in der Lage waren den Fragebogen eigenständig auszufüllen oder ihn mit ihrem Bezugsbetreuer ausfüllten, waren die Voraussetzungen „Motivation“ und „Fähigkeit“ für den Einsatz eines Fragebogens als Erhebungsinstrument erfüllt.

Die Bedarfserhebung wurde in den stationären Einrichtungen und ambulanten Angeboten für Menschen mit erworbenen Hirnschädigungen in Dortmund, Witten und Düsseldorf durchgeführt. Bei den Probanden handelte es sich um erwachsene Menschen mit einer erworbenen Hirnschädigung. Es wurde keine Vorauswahl hinsichtlich der vorhandenen Fähigkeiten oder der beeinträchtigten Funktionen getroffen. Von den insgesamt 80 ausgewählten Personen konnten am Ende der Umfrage 30 Befragungen ausgewertet werden. Das Alter dieser 30 Probanden (11 weiblich, 19 männlich) lag zwischen 25 und 65 Jahre (m=51). Die bei den Probanden vorliegende Hirnschädigung wurde durch verschiedene Ereignisse verursacht und ereignete sich in den Jahren zwischen 1982 und 2014 (m=2010), lag also im Schnitt 4 bis 5 Jahre zurück.

2.5 Hypothesen

Mit Hilfe der Nutzerbedarfsanalyse sollten die aktuelle Informations- und Kommunikationssituation, die tatsächlich vorhandenen Bedürfnisse und die damit einhergehende Beurteilung der allgemeinen Lebensqualität von Nutzern mit einer erworbenen Hirnschädigung erfasst werden. Dazu wurden auf der Grundlage der statistischen Testtheorie Hypothesen zu den zu erwartenden Unterschieden bei der Nutzung und den Nutzungsbedürfnissen vor und nach dem Eintritt der Hirnschädigung bzw. im Vergleich zwischen den Nutzern mit und ohne Beeinträchtigungen formuliert. Da unterschiedliche Untersuchungen gezeigt haben, dass die Fähigkeiten in der Nutzung von Alltagstechnologie nach einer Hirnschädigung mehr oder weniger stark beeinträchtigt sind, war davon auszugehen, dass Unterschiede in den Nutzungswünschen und dem Nutzungsverhalten gefunden werden können ([GD15], [So14]). Um die Richtigkeit dieser Annahmen zu zeigen, wurde in den Nullhypothesen jeweils davon ausgegangen, dass keine Unterschiede vorliegen.

Für die Erfassung der aktuellen Kommunikationssituation wurde die folgende Nullhypothese aufgestellt:

H_0**:** Es gibt keinen Unterschied in der Nutzung moderner Kommunikationsmedien, d.h. Menschen mit erworbenen Hirnschädigungen nutzen die modernen Kommunikationsmedien genauso häufig und in der gleichen Weise wie Menschen ohne Beeinträchtigungen.

Für die bei der untersuchten Nutzergruppe tatsächlich vorliegenden Informations- und Kommunikationsbedürfnisse wurde die folgende Nullhypothese aufgestellt:

H_0**:** Es gibt keine Diskrepanz zwischen dem Nutzungswunsch und der tatsächlichen Nutzung, d.h. Menschen mit erworbenen Hirnschädigungen haben vor der Schädigung die Kommunikationsmedien „normal" genutzt und nutzen diese weiterhin in der gleichen Weise.

In Bezug auf die Lebenszufriedenheit in Abhängigkeit von den Teilhabemöglichkeiten wurde die folgende Nullhypothese aufgestellt:

H_0: Die eigenständige Nutzung moderner Kommunikationsmedien wirkt sich nicht auf die Zufriedenheit mit der allgemeinen Kommunikations- und Lebenssituation aus, d.h. Menschen, die Kommunikationsmedien erfolgreich nutzen sind mit ihrer Kommunikations- und Lebenssituation nicht zufriedener als Menschen, die diese Kommunikationsmedien nicht nutzen können.

Die Nutzerbedarfsanalyse wurde als quasi-experimentelle Feldstudie angelegt, um eine möglichst hohe externe Validität zu erreichen. Bei der Auswahl der Probanden wurde durch Überprüfung bzw. durch Befragung der Bezugsbetreuer sichergestellt, dass die Probanden ein inhärentes Interesse an der Studienteilnahme haben und in der Lage sind, die Fragen zu verstehen und wahrheitsgemäß zu beantworten. Da die Studie auf der Basis der Ergebnisse der Erhebung über die Nutzung von Informations- und Kommunikationstechnologien konzipiert wurde, ist die Möglichkeit für einen direkten Vergleich mit einer altersentsprechenden Gruppe von Nutzern ohne Beeinträchtigungen gegeben, so dass im Rahmen der Feldstudie auf die Befragung einer Kontrollgruppe verzichtet werden konnte.

2.6 Ergebnisse

Für die Bereiche der aktuellen Nutzung und der Nutzungsbedürfnisse wurden verschiedene Typen von Endgeräten (Telefon, Handy, PC (Desktop, Laptop, Tablet)) und verschiedene Kommunikations- und Informationsprozesse unterschieden (allgemeine Internetnutzung, Email, Videotelefonie, Informationssuche, Unterhaltung, eCommerce) unterschieden.

Die Nutzung der Endgeräte hatte sich für alle drei Typen der Endgeräte nach dem Eintritt der Hirnschädigung deutlich verändert. Etwa ein Viertel der Probanden kann die vorher genutzten Endgeräte in der aktuellen Lebenssituation nicht mehr oder nur noch eingeschränkt nutzen (siehe Tabelle 1). 8 von 29 Probanden nutzen das Telefon aktuell nicht mehr, obwohl sie es gern nutzen würden, ebenso wird das Handy nur von 17 Personen genutzt, obwohl 21 der befragten Probanden einen entsprechenden Nutzungswunsch haben. Alle befragten Probanden hatten zwar deutliche körperliche und kognitive Beeinträchtigungen, waren aber grundsätzlich in der Lage zumindest ein Telefon oder ein Handy bedienen zu können.

Endgerät	Nutzung aktuell	Nutzung vorher	Nutzungswunsch
Telefon			
Ja	21	29	29
Nein	9	1	0
Handy			
Ja	17	20	21
Nein	11	10	0
PC			
Ja	15	20	24
Nein	15	10	0

Tab. 1: Endgerätenutzung vor und nach Eintritt der Hirnschädigung

Bezüglich der Nutzung der Kommunikations-, Informations- und Unterhaltungsangebote vor und nach Eintritt der Hirnschädigung konnten ähnliche Ergebnisse gefunden werden. Auch hier nutzt ein erheblicher Teil der Probanden die vorher genutzten Angebote nach dem Eintritt der Hirnschädigung nicht mehr (siehe Tabelle 2).

Wie ein Vergleich mit der altersentsprechenden Gruppe aus der IKT-Erhebung [DES15] zeigt, dass vor allem Email-Programme und Informationsdienste häufig genutzt werden. Diese Nutzungswünsche lassen sich auch bei den im Rahmen dieser Studie befragten Probanden als die häufigsten Nutzungswünsche finden. Die Nutzung von Videotelefonie und eCommerce-Diensten sowie die Teilhabe an sozialen Netzwerken und das Nutzen von Online-Spielen wird eher weniger gewünscht, das ist aber möglicherweise durch die Altersstruktur der befragten Probanden begünstigt. Untersuchungen von Patientengruppen unterschiedlicher Alterskohorten mit Hemiparesen haben gezeigt, dass jüngere Probanden ein großes Interesse an der Teilhabe an sozialen Netzwerken und Online-Spielen haben [La15].

Die kumulierten Daten zur Zufriedenheit mit der bestehenden Kommunikations- und Informationssituation sind in Tabelle 3 zu finden. Ein großer Teil der Probanden ist mit der bestehenden Situation eher zufrieden als unzufrieden, wobei sich die Unzufriedenheit im Bereich der Kommunikation eher auf das Verhalten der Kommunikationspartner als auf die zur Kommunikation verwendeten Medien bezieht. Die Probanden, die in der Lage sind die Kommunikationsmedien erfolgreich zu nutzen, sind mit ihrer allgemeinen Kommunikations- und Lebenssituation eher sehr zufriedenoder zufrieden während die Probanden, die die Medien nicht nutzen oder nicht nutzen können eher zufrieden oder unzufrieden sind.

Auf die Frage nach Änderungswünschen bezüglich der Kommunikationssituation wünschen aber dennoch 13 der Probanden deutliche Veränderungen. Dabei werden neben mehr Akzeptanz und Geduld bei den Kommunikationspartnern auch eine bessere Geräteausstattung, eine verbesserte Bedienbarkeit der Geräte und deutlich mehr Unterstützung beim Umgang mit Informations- und Kommunikationstechnologien gewünscht.

Kommunikationsprozess	Nutzung aktuell	Nutzung vorher	Nutzungswunsch
Email			
Ja	9	19	17
Nein	16	6	0
Videotelefonie			
Ja	3	4	12
Nein	22	21	0
Soziale Netzwerke			
Ja	3	6	10
Nein	22	19	0
eCommerce			
Ja	3	10	13
Nein	12	15	0
Recherche			
Ja	8	14	16
Nein	17	11	0
Unterhaltung			
Ja	3	9	13
Nein	22	16	0
Onlinespiele			
Ja	4	6	10
Nein	21	19	0
Nachrichten			
Ja	6	9	13
Nein	19	16	0

Tab. 2: Nutzung der Kommunikations-, Informations- und Unterhaltungsdienste vor und nach Eintritt der Hirnschädigung

3 Diskussion der Ergebnisse und Planung der weiteren Projektphasen

Im Rahmen der hier vorgestellten Nutzerbedarfsanalyse wurden drei Hypothesen zur Nutzungshäufigkeit von Informations- und Kommunikationstechnologien, zu den damit verbundenen Nutzungswünschen und zur allgemeinen Lebens- und Kommunikationszufriedenheit aufgestellt. Auch wenn die Datenbasis mit insgesamt 30 auswertbaren Fragebögen relativ gering war, lassen sich deutliche Tendenzen in den Ergebnissen erkennen.

Auf Basis der erhobenen Daten kann die Hypothese, dass die Nutzungshäufigkeit modernen Informations- und Kommunikationstechnologien nach dem Eintritt der Hirnschädigung unverändert zu vorher ist, als widerlegt gelten. Die befragten Probanden gaben deutliche Unterschiede im Nutzungsumfang der Kommunikationsmedien nach einer erworbenen Hirnschädigung im Vergleich zu vorher an.

Bereich	sehr zufrieden	eher zufrieden	eher unzufrieden
Geräte	6	18	6
Kommunikationspartner	3	16	11
Erreichbarkeit	3	20	7
Hilfestellung	10	15	5
Gesamtzufriedenheit	2	24	4
Zufriedenheit in Abhängigkeit von der Mediennutzung			
Ja	2	12	1
Nein	0	12	3

Tab. 3: Kommunikationszufriedenheit und allgemeine Zufriedenheit nach der Hirnschädigung

Auch die Hypothese, dass es keinen Unterschied zwischen der aktuellen Nutzung und dem vorhandenen Nutzungswunsch gibt, kann auf Basis der Ergebnisse als widerlegt gelten. Für die hier untersuchte Gruppe der Nutzer mit einer erworbenen Hirnschädigung konnten signifikante Unterschiede zwischen dem Nutzungswunsch und der tatsächlichen Nutzung nachgewiesen werden.

Die Antworten zur Zufriedenheit mit der allgemeinen Lebens- und Kommunikations-situation der hier befragten Probanden deuten darauf hin, dass eine erfolgreiche Mediennutzung positiv auf die Zufriedenheit auswirkt. Somit kann auch in diesem Bereich die Nullhypothese als widerlegt gelten.

Die wichtigsten Ergebnisse der detaillierten qualitativen Datenauswertung lassen sich wie folgt zusammenfassen:

- Menschen mit Beeinträchtigungen nach einer erworbenen Hirnschädigung im betreuten Wohnumfeld haben Bedürfnisse bei der Nutzung der modernen Kommunikationsmittel, die aktuell nicht erfüllt werden.
- Es werden konkrete Nutzungsvorstellungen geäußert und diese werden differenziert begründet.
- Überraschend häufig wird als Grund für die Nichtnutzung das Fehlen der Endgeräte genannt.
- Nur wenige der befragten Personen scheitern auf Grund ihrer kognitiven oder körperlichen Beeinträchtigungen an der Nutzung der Kommunikationsmittel. Die Bezugsbetreuer begründen die Nichtnutzung aber durch den z.T. sehr hohen Hilfebedarf der Probanden.
- Einige der befragten Personen nutzen die modernen Kommunikationsmittel überhaupt erst oder verstärkt nach dem Eintritt der Hirnschädigung.
- Die erfolgreiche Nutzung der modernen Kommunikations- und Informationsmittel hat einen positiven Einfluss auf die Lebens- und Kommunikationszufriedenheit.

Anders als erwartet konnte im Rahmen dieser Erhebung nicht nachgewiesen werden, dass Menschen mit Beeinträchtigungen nach einer erworbenen Hirnschädigung in erster Linie am Umgang mit den modernen Kommunikationsmitteln scheitern. Viele der befragten Personen, die die Kommunikationsmittel auch vor ihrer Erkrankung genutzt haben, gaben an, dass sie die Geräte mit veränderten Grundeinstellungen weiterhin nutzen können. Ein erweiterter Bedarf für spezielle Gerätesoftware oder modifizierte Nutzerschnittstellen wurde von der befragten Nutzergruppe nicht angegeben. Allerdings lässt der von den Betreuern angemerkte z.T. erhebliche Hilfebedarf bei der Nutzung darauf schließen, dass vor allem die Personen von vereinfachten Nutzerschnittstellen profitieren würden, die die modernen Kommunikationsmittel aktuell nicht nutzen können. Auch wenn es sich bei der Gruppe der Nutzer mit erworbenen Hirnschädigungen um eine Gruppe mit einem sehr heterogenen Beeinträchtigungsbild handelt, sind bestimmte kognitive Fähigkeiten, wie z.B. die Gedächtnis- und Konzentrationsleistungen bei fast allen Nutzern beeinträchtigt. Das lässt vermuten, dass die meisten Nutzer mit einer erworbenen Hirnschädigung Probleme bei der Nutzung komplizierter Nutzerschnittstellen, der Informationssuche auf inhaltlich und designtechnisch komplexen Internetseiten und den Zeitbeschränkungen, die bei der Ausführung bestimmter Funktionen auftreten, haben. Diese Annahmen decken sich mit den Ergebnissen vorhandener Studien (z.B. [Ge13]) und müssen, auch wenn sie in der hier vorgestellten Studie nicht explizit benannt worden sind, bei der Konzeption einer assistiven Arbeitsumgebung bedacht werden. Die Einschränkung der Probandengruppe auf Menschen, die in einem betreuten Wohnumfeld leben und in der Regel unter schwereren Beeinträchtigungen leiden als eine eigenständig lebende Vergleichsgruppe, kann dazu geführt haben, dass bestimmte Nutzungswünsche in der vorliegenden Studie nicht abgebildet oder nicht ihrem auf die Grundgesamtheit zutreffenden Umfang berücksichtigt worden sind.

Insgesamt sind sich die Autoren der Tatsache bewusst, dass sowohl das Design der Nutzerbedarfsanalyse als auch die Datensammlung Limitationen aufweisen, die eine umsichtige Interpretation der Ergebnisse erforderlich machen. Da die Nutzerbedarfsanalyse als Feldbefragung ohne Kontrollgruppe nur an einer kleinen Probandengruppe durchgeführt wurde, liegt eine eingeschränkte interne und externe Validität der Daten vor, die bei der Interpretation der Ergebnisse berücksichtigt wurde.

Auf Basis der Ergebnisse der Nutzerbedarfsanalyse werden für den weiteren Projektverlauf die folgenden Aufgaben angedacht:

- Konzeption einer integrierten Unterstützungsumgebung für die Nutzung der Endgeräte und der entsprechenden Anwendungen für die untersuchten Nutzergruppe
- Entwicklung einer nutzergruppengerechten Schnittstelle für webbasierte Kommunikationssoftware
- Detaillierte Untersuchung der kompensatorischen Einsatzmöglichkeiten moderner Kommunikationsmittel

Die prototypische Entwicklung einer integrierten Unterstützungsumgebung erscheint derzeit als die vielversprechendste Weiterentwicklung des Projektes. Mit Hilfe der Projekt-

partner und insbesondere mit Unterstützung des Projektes PIKSL der gGmbH In der Gemeinde leben Düsseldorf soll ein Prototyp einer integrierten Unterstützungsumgebung für die selbständige Nutzung von Kommunikationsgeräten und -anwendungen konzipiert und entwickelt werden. Entsprechend der technischen Realisierungsmöglichkeiten besteht die Unterstützungsumgebung entweder aus einem leistungsfähigen PC sowie diversen Eingabe- und Ausgabegeräten mit zusätzlichen Schnittstellen für die Integration weiterer Endgeräte oder einem serverbasierten Anwendungspaket, welches die Nutzer auf verschiedenen Endgeräten mit ihren Bedürfnissen angepassten Konfigurationseinstellungen nutzen können. Das Nutzerprofil und die Konfigurationseinstellungen werden in den Grundeinstellungen manuell angelegt und umfassen Informationen zu den präferierten Ein- und Ausgabekanälen, den Zugriffs- und Antwortgeschwindigkeiten sowie Listen häufig genutzter Anwendungen und Kontakte. Abbildung 3 zeigt das Grundkonzept der integrierten Unterstützungsumgebung.



Abb. 3: Grundkonzept einer integrierten Unterstützungsumgebung

Das Nutzerprofil wird auf Basis der von den Nutzern durchgeführten Aktionen ständig aktualisiert und modifiziert, so dass die Profilierung im Laufe der Zeit immer besser auf die tatsächlichen Fähigkeiten des Nutzers abgestimmt werden kann und zu erwartende Leistungsverbesserungen im Umgang mit der Technologie auch in veränderten Nutzereinstellungen abgebildet werden kann.

Die Entscheidung über den konkreten Einsatz von informationstechnischen Verfahren für die Profilierung wird im Rahmen der Konzeption gefällt. Geplant ist auch die Einbindung bzw. Entwicklung einer Anwendung, die den Nutzern gegenseitige Hilfestellungen im Rahmen eines Community-Ansatzes ermöglicht. In Abhängigkeit von den technischen Voraussetzungen könnten im weiteren Verlauf auch die Funktionen wie die elektronische Terminplanung, die Nutzung von webbasierten Therapieangeboten oder eLearning-Aktivitäten in die vorhandene Infrastruktur eingebunden werden.

Da die meisten Einrichtungen des betreuten Wohnens technisch nicht entsprechend ausgestattet sind, muss im Rahmen dieses Projektes die Beschaffung einer infrastrukturellen Grundausstattung geprüft werden. Um eine erfolgreiche, über die Dauer des Projektes hinausgehende Nutzung des Arbeitsplatzes und der entwickelten Applikationspakete gewährleisten zu können, müssen die Einrichtung, Administration und Kontrolle von einem der Projektpartner oder einem externen Dienstleister dauerhaft übernommen werden, da die Einrichtungen dieses erfahrungsgemäß nicht alleine bewältigen können.

4 Danksagung

Diese Arbeit wurde durch die Unterstützung der Stiftung Bethel, Bethel.regional und die gGmbH „In der Gemeinde Leben“ Düsseldorf ermöglicht.

Literaturverzeichnis

[Bu14] Bundesministerium für Arbeit und Soziales: Leichte Sprache. Ein Ratgeber. Bericht, http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/a752-ratgeber-leichte-sprache.pdf [24.02.2015], 2014.

[De05] Deutsches Institut für Medizinische Information und Dokumentation: Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit. Bericht, WHO, Genf, 2005.

[DES15] DESTATIS. Private Haushalte in der Informationsgesellschaft (IKT), Fachserie 15, Reihe 4, https://www.destatis.de/DE/Publikationen/Thematisch/EinkommenKonsumLebensbedingungen/PrivateHaushalte/PrivateHaushalteIKT.html [10.06.2015], 2015. Statistisches Bundesamt.

[GD15] Goverover, Y.; DeLuca, J.: Using the Internet to assess everyday functioning after traumatic brain injury. In: Brain Injury. S. 1–7, 2015.

[Ge13] Gell, N.; Rosenberg, D.; Demiris, G.; LaCroix, A.; Patel, K.: Patterns of Technology Use Among Older Adults With and Without Disabilities. In: The Gerontologist. 2013. doi: 10.1093/geront/gnt166.

[HFE09] Huenerfauth, M.; Feng, L.; Elhadad, N.: Comparing Evaluation Techniques for Text Readability Software for Adults with Intellectual Disabilities. In: Assets ’09 Preceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility. S. 3–10, 2009.

[KPLL13] Kassberg, A.; Prellwitz, M.; Larsson Lund, M.: The challenges of everyday technology in the workplace for persons with acquired brain injury. In: Skandinavian Journal of Occupational Therapy. S. 272–281, 2013.

[La15] Lam, M.Y.; Tatla, S.K.; Lohse, K.R.; Shirzad, N.; Hoens, A.M.; Miller, K.J.; Holsti, L.; Virji-Babul, N.; Van der Loos, M.: Perceptions of Technology and its Use for Therapeutic Application for Individuals with Hemiparesis: Findings from Adult and Pediatric Focus Groups. In: LMIR Rehabilitation and Assistive Technologies. Jgg. 2(1), 2015.

[MKNa13] Malinowsky, C.; Kottorp, A.; Nyg ˚ ard, L.: Everyday technologies' level of challenge when used older adults with and without cognitive impairment - comparision of self-perceived versus observed difficulty estimates. In: Technology and Disability. Jgg. 25(3)), S. 167–176, 2013. doi: 10.3233/TAD-130380.

[So14] Soria, E.; Joseph, P.A.; Kaoua, B.N.; Simion, A.; Mazaux, J.M.; Klinger, E.: Performance analysis of adults with Aquired Brain Injury making errands in a virtual supermarket. In: Proceedings of the 10th Conference of Disability, Virtual Reality & Associated Technologies. Gothenburg, Schweden, 2014.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Erweiterung eines virtuellen Klassenzimmers zur Verbesserung der Zugänglichkeit für Blinde

Wiebke Köhlmann[1], Nils Dressel[1], Dustin Wegner[1]

Abstract: Virtuelle Klassenzimmer ermöglichen die Durchführung von Veranstaltungen in Echtzeit an verteilten Standorten. Die Teilnahme von blinden Lernenden wird jedoch durch vielfältige Barrieren behindert. Dieser Beitrag beschreibt eine prototypische Erweiterung des virtuellen opensource Klassenzimmers BigBlueButton um ein Aktivitätsprotokoll, Konfigurationsmöglichkeiten, eine Notiz- und Beschreibungs-Funktion sowie die Verbesserung der Screenreader-Kompatibilität und Aufzeichnungswiedergabe mit dem Ziel der Verringerung von Zugänglichkeitsbarrieren für Blinde.

Keywords: Barrierefreiheit, Inklusion, Menschen mit Beeinträchtigung, blind, virtuelle Klassenzimmer, kollaboratives Lernen, E-Learning, Mensch-Maschine-Interaktion

1 Einleitung

Virtuelle Klassenzimmer sind synchrone Konferenzsysteme zur Durchführung von Veranstaltungen in Echtzeit an verteilten Standorten [FH12]. Sie können viele Eigenschaften einer Präsenzlehr- bzw. -lernsituation mit Hilfe von synchroner Kommunikation und Medieneinsatz (bspw. mit Chat, Audio- und Videokonferenz und dynamischem Whiteboard) abbilden. Inklusion und die Verwendung interaktiver und kollaborativer E-Learning-Werkzeuge gewinnen in Bildungsbereichen an Bedeutung. Jedoch entstehen bei der Verwendung von Anwendungen, wie bspw. virtuellen Klassenzimmern, neue Barrieren für Lernende mit Beeinträchtigungen aufgrund zunehmender Verwendung grafischer, dynamischer und synchroner Inhalte und Bedienoberflächen. Der Zugang zu Information erfolgt für blinde Lernende mittels technischer Hilfsmittel, meistens mit Sprachausgabe über einen Screenreader und taktile Ein-/Ausgabegeräte, sogenannte Braille-Zeilen. Damit erfolgt die Ausgabe textbasiert und linear – grafische Darstellungen, dynamische visuelle Medien, strukturelle und räumliche Zusammenhänge sowie parallele Ereignisse können mit diesen Hilfsmitteln nicht abgebildet werden, auch wenn die Bedienung mittels Screenreader von der Anwendung u. a. durch Tastaturbefehle für alle Funktionen und einer logischen Tabulator-Reihenfolge unterstützt wird.

Nachfolgend wird die Barrierefreiheit virtueller Klassenzimmer beschrieben und das virtuelle Klassenzimmer BigBlueButton [BBB15] vorgestellt. Darauf werden die vorgenommenen Erweiterungen von BigBlueButton um alternative Konzepte zur Verbesserung der Barrierefreiheit beschrieben.

[1] Universität Potsdam, Institut für Informatik und Computational Science, August-Bebel-Str. 89, 14482 Potsdam, wiebke.koehlmann@uni-potsdam.de, ndressel@uni-potsdam.de, dustin.wegner@uni-potsdam.de

2 Barrierefreiheit virtueller Klassenzimmer

2.1 Richtlinien

Für die barrierefreie Gestaltung von Webseiten und Anwendungen existieren verschiedene Richtlinien mit unterschiedlichen Ausprägungen bspw. in Bezug auf den Typ der Zielanwendungen. Empfehlungen aus verschiedenen Richtlinien (u. a. [DIN06, IMS04, W3C08]) ergänzen sich in Hinblick auf relevante Zugänglichkeitsaspekte für virtuelle Klassenzimmer. Erweitert durch weitere Gestaltungshinweise ergeben sich daraus konkrete Richtlinien für die barrierefreie Gestaltung virtueller Klassenzimmer [Kö14]. Nachfolgend werden die wichtigsten Richtlinien und Maßnahmen zusammengefasst:

1) Wahrnehmung: Alle Inhalte müssen für alle Nutzer wahrnehmbar und vorhersehbar sein. Für alle Nicht-Text-Objekte müssen Alternativen oder Beschreibungsmechanismen angeboten werden und die Wahrnehmung von Änderungen muss ermöglicht werden.

2) Navigation: Alle Mausaktionen müssen auch mit der Tastatur ausführbar sein. Zum Ansteuern von und Wechseln zu wichtigen Elementen müssen Tastaturbefehle verfügbar sein. Die Navigation in zeitbasierten Medien ist mittels Zeitleiste zu ermöglichen.

3) Orientierung: Die Bedienoberfläche soll einer klaren und logischen Struktur folgen und Orientierungshilfen bieten. Mechanismen zum Erkennen von Beziehungen und Abhängigkeiten, bspw. mittels Hyperlinks, sollten verfügbar sein.

4) Interaktion: Alle Nutzenden müssen gleichberechtigt teilhaben können. Eine (Eingabe-)Unterstützung mittels kontextsensitiver Rückmeldungen sollte angeboten werden.

5) Semantik: Die Abhängigkeiten und Verbindungen zwischen Inhaltselementen müssen erkennbar und die Autoren von Beiträgen müssen identifizierbar sein.

6) Geschwindigkeit: Teilnehmende sollen in der Lage sein die Veranstaltung und Inhalte in ihrer eigenen Geschwindigkeit mit eigenen technischen Hilfsmitteln zu wiederholen. Dafür muss eine individuelle Kontrolle zeitbasierter Medien gewährleistet werden.

7) Soziale Präsenz: Der eigene Teilnehmerstatus und der Anderer muss erkennbar sein.

8) Konfiguration: Zur Anpassung der Anwendung an persönliche Bedürfnisse sollten profilbasierte Anpassungen der Bedienoberfläche, Inhaltspräsentation, Interaktionsmodalitäten und Benachrichtigungsmechanismen verfügbar sein.

9) Hilfestellung: Eine Unterstützung bei Eingaben und der Orientierung sollte vorhanden sein. Alle Teilnehmenden sollten in der Lage sein, anderen Teilnehmenden Hilfestellung zu leisten und Inhalte anderer zu kontrollieren.

2.2 Verwandte Arbeiten

Untersuchungen zu der Barrierefreiheit virtueller Klassenzimmer sind eher selten. Daher werden nachfolgend auch Anwendungen mit verwandter Funktionalität zu virtuellen Klassenzimmern beschrieben.

Um ein dynamisches Whiteboard in einem virtuellen Klassenzimmer zugänglich zu gestalten, integrieren [Fr10] einen menschlichen Übersetzer, welcher alle nicht-visuellen Inhalte in textueller Form in Echtzeit beschreibt. Dies bedarf einer umfassenden Vorbereitung von Lehrkraft und Übersetzendem und ist bei häufigem Einsatz sehr kostenintensiv. [SCM14] analysieren die Zugänglichkeit von virtuellen Klassenzimmern und entwickeln weiterhin ein kollaboratives Whiteboard mit grundlegenden Funktionen unter Berücksichtigung gängiger Barrierefreiheits-Richtlinien. Die Plattform unterstützt u.a. Alternativtexte für Nicht-Text-Elemente, ARIA-Attribute, verschiedene Geräte (Maus, Tastatur u.a.) zur Verwendung der Whiteboard-Funktionen, Tastaturkürzel, Tabulator-Index, Screenreader-Unterstützung für das Chat-Werkzeug, das Bewegen von editierbaren Elementen mittels Tastatur und auditive Benachrichtigungen über Änderungen.

Barrieren in Bezug auf Informationsüberlastung, eine Vermischung von Inhalt und Struktur wie am Beispiel von Google Docs beschrieben [Mo11], treten auch in virtuellen Klassenzimmern auf. Zur besseren Verwendbarkeit durch blinde Nutzende wurde die Tastatur- und Screenreader-Unterstützung von Google Docs durch [Mo11] verbessert, so dass die Werkzeug-Elemente bedienbar sind und Texteigenschaften auch mittels Tastaturkürzeln geändert werden können. Sowohl für Google Docs als auch für ein Wiki-basiertes System für Menschen mit Beeinträchtigungen [Me11] wurde ein zugänglicher WYSIWYG-Editor integriert.

2.3 Maßnahmen der Hersteller

Um den Zugang für blinde Teilnehmende an Veranstaltungen in virtuellen Klassenzimmern zu erleichtern, integrieren Hersteller einfache Barrierefreiheitsfunktionen in ihre Lösungen. Diese umfassen bspw. eine vollständige Tastaturunterstützung mit logischer Tabulator-Reihenfolge und Tastaturkürzeln, Audio-Benachrichtigungen und Screenreader-Unterstützung. Einige virtuelle Klassenzimmer bieten weitere Funktionen an, bspw.:

- Ein Aktivitätsfenster dokumentiert die Ereignisse einer Sitzung mit ihren Zeitstempeln und Inhalten [Bl14].
- Eine Beschreibungsfunktion ermöglicht die Eingabe von Alternativtexten für dargestellten Inhalt [Bl14, TC08].
- Zugängliche Aufzeichnungen ermöglichen es Teilnehmenden die Aufzeichnungen in ihrem eigenen Tempo zu wiederholen [Bl14].
- Konfigurierbare Audiobenachrichtigungen und Tastaturkürzel ermöglichen Anpassung der Bedienoberfläche an die Bedarfe der Teilnehmenden [Bl14, TC08].

2.4 Alternative Konzepte für barrierefreie virtuelle Klassenzimmer

Allein auditive Benachrichtigungen über Ereignisse (vgl. [SCM14]) sind nicht ausreichend, um alle Änderungen wahrnehmen zu können. Die Ergänzung um ein Aktivitätsprotokoll (vgl. [Bl14]) unterstützt das Verfolgen des Sitzungsablaufs auch für langsam kommunizierende Nutzende. Trotz der Möglichkeit der Bedienbarkeit des Whiteboards mit der Tastatur (vgl. [SCM14]), sind semantische Zusammenhänge der Elemente nur schwer erkennbar. Weiterhin hat ein blinder Nutzender zwar die Möglichkeit, Whiteboard-Elemente zu editieren, aber nicht diese zu erstellen. Um alle Elemente auf dem Whiteboard erfassen zu können, sind alternative Beschreibungen erforderlich, aber nicht immer sind zusätzliche menschliche Ressourcen wie bei [Fr10] dazu verfügbar.

Bestehende Ansätze zur Verbesserung der Barrierefreiheit virtueller Klassenzimmer – insbesondere für blinde Teilnehmende – sind wichtig, jedoch erfüllen diese nicht alle relevanten Richtlinien [Kö14]. Daher wurden auf dieser Basis weiterführende alternative Bedienkonzepte für virtuelle Klassenzimmer entwickelt und durch Benutzertests mit blinden Probanden mittels taktilem Paper-Prototyping bestätigt [KL15]. Diese umfassen, neben grundlegenden Anpassungen wie Tastaturunterstützung und Screenreader-Kompatibilität, u. a. eine Beschreibungsfunktion für Objekte auf dem Whiteboard, die Möglichkeit verwandte Inhalte zu verlinken (bspw. eine Frage im Chat mit dem entsprechenden Element auf dem Whiteboard) und Lesezeichen zu setzen, ein Aktivitätsprotokoll zur linearen Dokumentation aller Ereignisse im virtuellen Klassenzimmer mit Zeitstempel und Möglichkeiten zur Interaktion mit dem Whiteboard durch blinde Teilnehmende.

3 BigBlueButton

BigBlueButton [BBB15] ist ein virtuelles Klassenzimmer unter der Lizenz *GNU Lesser General Public License (LGPL)* [LGPL15]. Es unterstützt die Funktionen Chat, Audio- und Videokonferenz, Whiteboard, Desktop-Sharing, Teilnehmendenliste und Aufzeichnungen. Analysen bezüglich der Richtlinienkonformität virtueller Klassenzimmer [SK13] und bezüglich der Bedienbarkeit mit Screenreadern [Kö15] zeigen, dass BigBlueButton das zugänglichste Open-Source-Klassenzimmer unter den untersuchten Lösungen ist. Da eigene Konzepte zur Verbesserung der Barrierefreiheit integriert und die Ausgabe für verschiedene Ausgabegeräte ermöglicht werden sollen, war die Verwendung einer Open-Source-Lösung naheliegend.

3.1 Barrierefreiheit von BigBlueButton

Laut der Anleitung von BigBlueButton [BBB13] sind Barrierefreiheitsfunktionen für das Chat-Modul (Tastaturnavigation, Eingabefeldfokussierung, Screenreader-Kompatibilität und Audiobenachrichtigungen), die Aufzeichnung und Wiedergabe (Zugang zu Texten von Präsentationen und des Chats) und allgemeine Anpassungsmöglichkeiten (skalierba-

re Fenster, Fenster mit Befehlsliste der Tastaturkürzel sowie Tastaturbefehle für Lokalisierung) vorhanden. Weiterhin existieren Richtlinien für Entwickelnde, die bei der Erweiterung von BigBlueButton beachtet werden müssen [BBB14]: logische Tabulator-Reihenfolge, Tastaturkürzel und Screenreader-Kompatibilität (JAWS[2] und NVDA[3]) über das Setzen eines `accessibilityName` im ActionScript[4].

BigBlueButton wurde in der Version 0.9.0-beta verwendet, da diese Version ein aktuelles Betriebssystem unterstützt. Die stabile Version wurde im April 2015 nach Abschluss der nachfolgend beschriebenen Implementierungen veröffentlicht. Praktische Tests mit der Version 0.9.0-beta zeigten, dass eine Bedienung per Screenreader – entgegen der Herstellerangaben – nur eingeschränkt möglich war, da die Tabulator-Reihenfolge nicht immer konsistent und nicht alle Tastaturkürzel funktional waren [SK15]. Ein Auslesen der textuellen Folieninhalte ist in Version 0.9.0-beta nicht möglich und das Zeichnen auf dem Whiteboard kann von Nutzenden nicht wahrgenommen werden. Weiterhin führt der erforderliche Fensterwechsel zum Erfassen paralleler Aktionen zu einem Zeitverlust bei ohnehin verlangsamter Informationsaufnahme.

3.2 Bedienoberfläche

BigBlueButton unterstützt verschiedene Layouts zur Darstellung der Bedienoberfläche. Nachfolgend wird die Standard-Anordnung beschrieben (siehe Abb. 1). Im Zentrum der

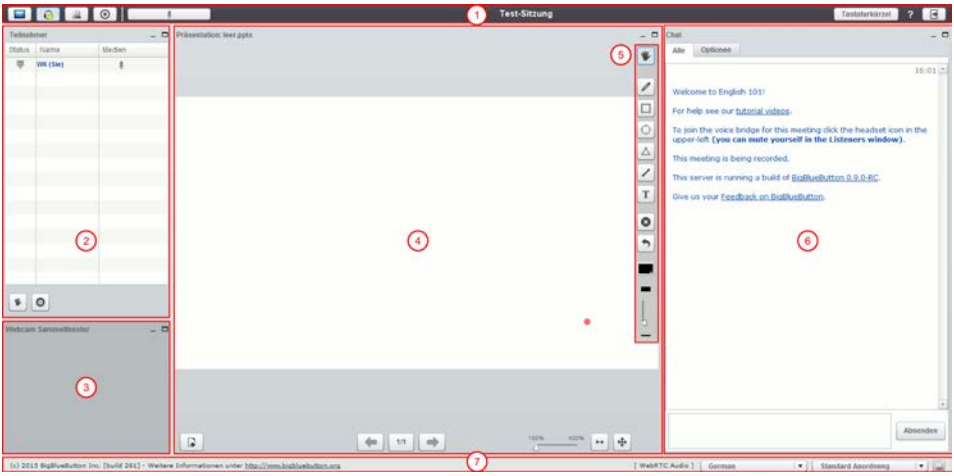


Abb. 1: Bedienoberfläche BigBlueButton v. 0.9.0, Präsentatoren-Ansicht. 1 Titelleiste, 2 Teilnehmendenliste, 3 Video, 4 Präsentationsbereich, 5 Werkzeugleiste, 6 Chat, 7 Einstellungen

[2] JAWS ist ein Screenreader von Freedom Scientific, http://www.freedomsci.de/prod01.htm
[3] NVDA ist ein freier Screenreader von NV Access, http://www.nvaccess.org/
[4] ActionScript ist eine Programmiersprache von Adobe Systems für bspw. Adobe Flash oder Flex

Bedienoberfläche von BigBlueButton steht das Whiteboard (4) u. a. zum Anzeigen von Präsentationsfolien und Erstellen von Zeichnungen und Anmerkungen. Im Falle von Präsentatoren-Rechten werden zur Steuerung der Folien darunter Schaltflächen zur Navigation angezeigt und bei Mausbewegungen auf dem Whiteboard wird eine Werkzeugleiste (5) eingeblendet. Auf der linken Seite werden die Liste aller Teilnehmenden (2) und darunter vorhandene Videoübertragungen (3) angezeigt. Auf der rechten Seite ist das Chat-Fenster (6) angeordnet. Über allen Bereichen ist eine Leiste mit Schaltflächen (1) zum (De-)Aktivieren von Shared-Desktop-Übertragungen, dem Mikrofon, der Kamera und der Aufzeichnung sowie der Titel der Veranstaltung und rechts Schaltflächen zu einer Tastaturkürzelübersicht, zur Hilfe und zum Verlassen der Sitzung angeordnet. Unten neben den Urheberrechtsinformationen (7) können Sprach- und Layout-Einstellungen vorgenommen werden.

3.3 Systemarchitektur

Die Architektur von BigBlueButton besteht aus mehreren gekapselten Komponenten. Im Folgenden sind die Hauptkomponenten des Systems kurz beschrieben:

- *bigbluebutton-apps*: serverseitige red5 und web-apps (Java/Scala)
- *bigbluebutton-client*: Flash-/Flex-Client (MXML/ActionScript)
- *bigbluebutton-web*: Grails-App für Konferenzfunktionen und Logging (Java)
- *deskshare-app*: serverseitiges Desktop-Sharing (red5 web-app) (Java)
- *deskshare-applet*: Applet zum clientseitigen Screencapturing (Java)

In Abb. 2 ist die Verbindung zwischen Client und den BigBlueButton-Applikationen erkennbar. Alle dargestellten Applikationen sind serverseitige Module, die mit den jeweilig korrespondierenden Client-Modulen kommunizieren. Die Client-Module sind in dieser Abbildung jedoch nicht explizit abgebildet, sondern im Knoten *client* zusammen-

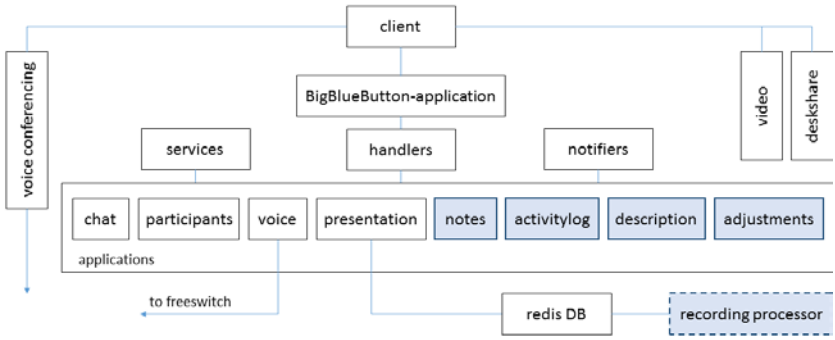


Abb. 2: Ausschnitt der Gesamt-Architektur: BigBlueButton-Applikationen (nach [BBB15]) mit neuen Modulen (blau schattiert) und angepasster Komponente für die Aufzeichnung (gestrichelt)

gefasst. Zur Anpassung und Erweiterung von BigBlueButton werden neue Module hinzugefügt und auf Serverkomponenten (*client*-Knoten) zugegriffen. Die Kommunikation für das Aktivitätsprotokoll erfolgt über die *services*, die Konfiguration greift auf die *notifiers* zu. Anpassungen zur Verbesserung der Screenreader-Kompatibilität erfolgen in den einzelnen Modulen.

4 Erweiterungen des virtuellen Klassenzimmers BigBlueButton

Nachfolgend werden von uns konzipierte und umgesetzte Erweiterungen des virtuellen Klassenzimmers BigBlueButton beschrieben. Dazu zählen die allgemeine Verbesserung der Barrierefreiheit und Screenreader-Kompatibilität, ein Aktivitätsprotokoll, eine Notizfunktion, eine Beschreibungsmöglichkeit für Whiteboard-Elemente und die Erweiterung der Aufzeichnungswiedergabe. Für die Anpassungen wurde aufgrund der Verbreitung und um eine Weiterentwicklung zu ermöglichen, Englisch als Sprache gewählt. Unterstützte Browser sind Firefox und Google Chrome, da bspw. der Internet Explorer die Tabulator-Reihenfolge nicht korrekt wiedergibt und die im .ogg[5]-Format gespeicherten Audioaufnahmen der Aufzeichnung nicht unterstützt.

4.1 Screenreader-Kompatibilität

Um eine grundlegende Bedienbarkeit mit Hilfe der Tastatur zu gewährleisten, müssen alle Bedienelemente mittels Tastatur erreichbar sein. Um Blinden die Bedienung zu ermöglichen, müssen zusätzlich Informationen zu der Funktionalität der Bedienelemente verfügbar sein, da ohne das Wissen über diese Funktionalität das reine Ansteuern mittels Tastatur keinen Mehrwert bietet. Neben einer logischen Tabulator-Reihenfolge durch alle Bedienelemente, sind Tastaturkürzel zum Ansteuern der wichtigsten Bedienelemente für eine schnelle Navigation hilfreich.

Tastaturkürzel

BigBlueButton bietet Tastaturkürzel zur Navigation an. Eine Liste der verfügbaren Befehle kann in einem zusätzlichen Fenster eingeblendet werden. Tests ergaben jedoch, dass die Tastaturkürzel nur teilweise funktionstüchtig sind und nicht alle Elemente mittels der Tabulator-Taste angesteuert werden können. In der verwendeten Beta-Version musste zunächst die Unterstützung der vorhandenen Tastaturkürzel wiederhergestellt werden. Bei der Wahl von Tastaturkürzeln ist bei Browser-basierten Anwendungen darauf zu achten, dass keine Konflikte mit den Tastaturkürzeln des Browsers auftreten und somit Befehle an das virtuelle Klassenzimmer nicht weitergegeben werden. Da die Erstellung Browser-übergreifender Tastaturkürzel bei Weiterentwicklungen der Browser eine kontinuierliche Anpassung erfordern würde, wurde stattdessen eine benutzerspezifi-

[5] .ogg ist ein Containerformat für Multimediadateien und kann somit gleichzeitig Audio-, Video- sowie Textdaten enthalten.

sche Konfiguration angestrebt. Da BigBlueButton keine Nutzerprofile vorsieht, musste die Konfiguration auf Seiten des Clients erfolgen. Daher wurde eine Benutzeroberfläche mittels JavaScript und HTML erstellt, welche die Möglichkeit bietet bestehende Tastaturkombinationen mit eigenen zu ersetzen.

Tabulator-Reihenfolge

Die Tabulator-Reihenfolge wurde überarbeitet. Dafür wird zunächst die Reihenfolge der Module definiert, um einen Wechsel zwischen den Modulen zu ermöglichen. Innerhalb der Module wird eine eigene Reihenfolge definiert. Die Teilnehmenden sind damit in der Lage mittels der Tabulator-Taste durch alle Module zu navigieren (vgl. Abb. 5).

Beschreibungen für Bedienelemente

In ActionScript können über die Eigenschaft `accessibilityName` Beschreibungen für Bedienelemente angelegt werden. Für jedes Bedienelement in BigBlueButton wurde im Rahmen der Anpassung ein `accessibilityName` hinzugefügt. Die Unterstützung verschiedener Sprachen ist dabei über Lokalisierungsdateien möglich. Tab. 1 zeigt für eine Auswahl von Bedienelementen des Chat-Moduls den `accessibilityName` und die entsprechende Screenreader-Ausgabe. Bei dem hier dargestellten Beispiel Jaws erfolgt zunächst die Ansage des hinterlegten `accessibilityName` und dann die Bezeichnung des Bedienelements.

Benachrichtigungen über Ereignisse in BigBlueButton können über Audio-Signale erfolgen. In BigBlueButton sind bereits einige Audio-Signale integriert, wie bspw. für das Ereignis *Mikrofon aktivieren* („You are now unmuted“) und *Mikrofon deaktivieren* („You are now muted“). Über die Konfiguration können weitere Audio-Signale – Töne und Sprachaufnahmen – benutzerspezifisch für folgende Ereignisse eingestellt werden. Dazu zählen u. a. neue öffentliche oder private Chat-Nachrichten, ein Wechsel der Präsentationsfolie, Starten und Beenden der eigenen Video-Übertragung sowie das Betreten und Verlassen der Sitzung durch einen Teilnehmenden.

Bedienelement	`accessibilityName`	Jaws-Ausgabe
Minimize	Minimize the chat window	„Minimize the chat window Button“
Chat input	Chat message editing field	„Chat message editing field Type a Text“
Send message	Send chat message	„Send chat message Button“

Tab. 1: Auswahl von Chat-Bedienelementen mit `accessibilityName` und Jaws-Ausgabe

4.2 Aktivitätsprotokoll

Ähnlich dem Aktivitätsfenster von Blackboard Collaborate[6] wurde ein Aktivitätsprotokoll-Modul (Modul *activitylog*) für BigBlueButton implementiert. In diesem Modul

[6] Blackboard Collaborate ist ein virtuelles Klassenzimmer des Herstellers Blackboard Inc. http://www.blackboard.com/Platforms/Collaborate/Overview.aspx

werden alle relevanten Aktivitäten (Betreten und Verlassen der Sitzung, Handmeldungen, Chat-Nachrichten, Folienwechsel der Präsentation, Whiteboard-Aktivitäten usw.) chronologisch mit einem Zeitstempel aufgelistet. Da nicht alle Informationen für alle Teilnehmenden interessant sind bzw. ein großer Umfang an Einträgen nicht ausreichend schnell erfasst werden kann, ist es möglich die anzuzeigenden Ereignisse durch Filter auf den persönlichen Bedarf anzupassen. Um eine direkte Teilnahme an der Veranstaltung ohne einen Fokuswechsel auf andere Bereiche des virtuellen Klassenzimmers zu ermöglichen, können Befehle direkt über eine unter dem Protokoll angeordnete Kommandozeile im Modul gegeben werden (vgl. Tab. 2). Diese umfassen derzeit 26 Befehle, u.a. Abfragen zu Teilnehmenden und Befehle zum Erstellen von Chat-Nachrichten.

Abb. 3 zeigt das Ausgabefenster des Aktivitätsprotokolls, in dem beispielhaft ein paar Aktivitäten und Kommandozeilenabfragen protokolliert wurden, und einen Ausschnitt des Optionen-Reiters, welcher Einstellungen zur Schriftgröße, Filtereinstellungen und Speichermöglichkeit anbietet. Im ersten Eintrag des Protokolls findet ein *USER*-Event statt, welches besagt, dass der Nutzer *Bob* die Konferenz betreten hat. Der zweite Eintrag entsteht durch die Abfrage `whoisuser Bob` auf der Kommandozeile. Es folgen eine Handmeldung sowie ein Seitenwechsel. Die textuellen Inhalte der Folie („slide 1“) werden durch die Eingabe des Kommandos `read` als *QUERY*-Eintrag angezeigt. Die Formatierung der Folieninhalte bleibt dabei soweit wie möglich erhalten.

Kommando	Aktion	Ausgabe
`filter <tags>`	Filtert die Ausgabe	Anzeige aller Nachrichten mit den als Argument übergebenen Tags
`list users`	–	„All available Users: <namelist>“
`mute me`	Schaltet Mikrofon stumm	„You are muted now.“

Tab. 2: Auszug verfügbarer Kommandos im Aktivitätsprotokoll

4.3 Notizen

Das Notizmodul (Modul *notes*) ermöglicht es allen Teilnehmenden von BigBlueButton private Notizen bzw. Mitschriften der Veranstaltung anzufertigen, ohne dafür zu einer weiteren Anwendung wechseln zu müssen und somit den Anschluss in einer Echtzeit-Sitzung zu verpassen. Bei der Darstellung soll eine Listenansicht für einzelne Notizen die Navigierbarkeit verbessern und einen Überblick erleichtern. Für eine bessere Auffindbarkeit der Notizeinträge wurde ein Konzept zur Titelgenerierung entwickelt.

Weiterhin soll das Erstellen von Notizen mittels Spracheingabe – sei es als Audiodatei oder Texterfassung per Spracherkennung – ermöglicht werden, um die Eingabe zu erleichtern. Das Abspeichern erfolgt lokal über eine Schaltfläche oder nach einer ersten manuellen Speicherung automatisch. Die derzeit erfolgte prototypische Umsetzung erlaubt das Verfassen von Notizen und das Abspeichern dieser.

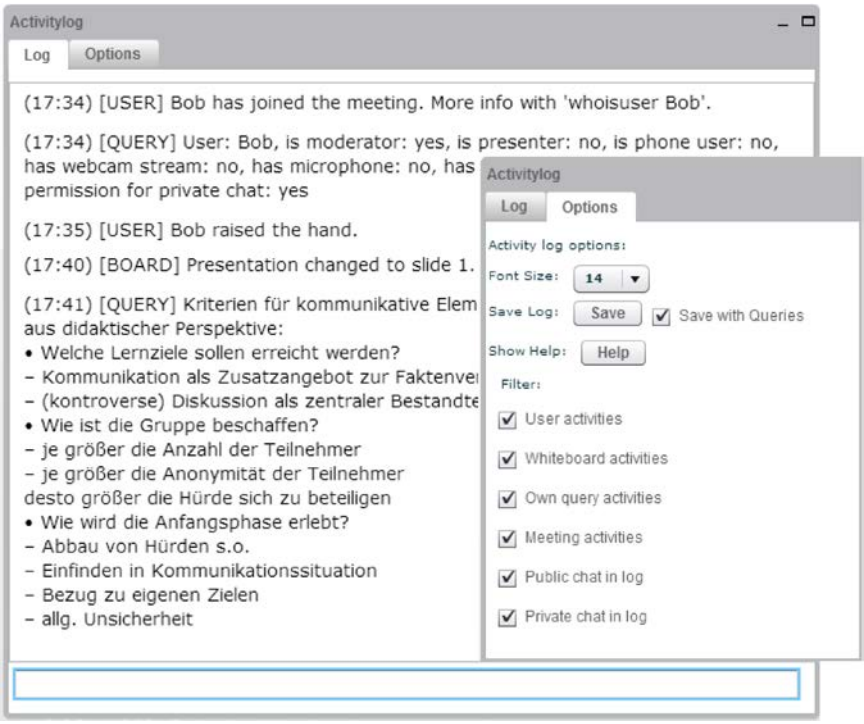


Abb. 3: Aktivitätsprotokoll und Optionen für Filter- und Anzeigeeinstellungen

4.4 Beschreibungen für Whiteboard-Elemente

Das Beschreibungsmodul für Whiteboard-Elemente (Modul *description*) listet alle gezeichneten Elemente einer Folie auf und erlaubt das kollaborative Hinzufügen einer Beschreibung für jedes Element. Dafür wird ein Element aus der Liste des Beschreibungsmoduls ausgewählt und auf dem Whiteboard mittels Rahmen gekennzeichnet. Über einen „Edit"-Button im Modul kann eine Beschreibung für das selektierte Element hinzugefügt werden. Solange ein Nutzender ein Element editiert, ist es für die anderen Teilnehmenden gesperrt. Es ist weiterhin möglich zu jedem Element Eigenschaften (bspw. Farbe und Position) und auch gelöschte Elemente in der Liste anzeigen zu lassen.

4.5 Aufzeichnung

Aufzeichnungen von Veranstaltungen in virtuellen Klassenzimmern ermöglichen es Teilnehmenden, die Veranstaltung in ihrer eigenen Geschwindigkeit und mit ihren eigenen Hilfsmitteln zu wiederholen. Somit ist auch die Barrierefreiheit der Aufzeichnung von großer Bedeutung.

Die Aufzeichnungen von BigBlueButton basieren auf HTML5, JavaScript und dem Mozilla Popcorn Framework[7] und unterscheiden sich somit von dem Echtzeit-Klassenzimmer in Bezug auf die Implementierung, Bedienoberfläche und Zugänglichkeit. Die Daten einer Veranstaltung in BigBlueButton werden in einer Datenbank gespeichert und mittels Popcorn geladen. Zur Verbesserung der Zugänglichkeit werden ARIA-Informationen hinterlegt.

Im Gegensatz zu dem Echtzeit-Klassenzimmer sind in der Aufzeichnung die Navigationsleiste und Folieninhalte mit einem Screenreader auslesbar. Die Bedienoberfläche der Veranstaltungs-Aufzeichnung von BigBlueButton besteht aus fünf Bereichen (vgl. Abb. 4): Überblick über alle Folien (1), aktuelle Folie (2), Reiter mit Einstellungen und öffentlichen Chat-Nachrichten (3) und Abspielkontrolle (4) sowie – falls vorhanden – Videoübertragung.

Zur Verbesserung der Barrierefreiheit wurden das Aktivitätsprotokoll und das Notiz-Modul in die Aufzeichnung integriert. Dazu erfolgte eine Anpassung der Bedienoberfläche: Das Chatfenster wurde dazu um die Reiter *Notes*, *Activitylog* und *Settings* ergänzt (vgl. Abb. 4). Im Aktivitätsprotokoll werden nur die in der Aufzeichnung dargestellten Inhalte dokumentiert. Dazu zählen öffentliche Chat-Nachrichten, Zeichnungen auf einer Folie, Texte auf der Folie, Folienwechsel, Wechsel des Vortragenden und Desktop-Sharing. Da BigBlueButton keine Nutzerauthentifizierung vorsieht, müssen private Notizen manuell geladen werden. Dafür ist vorgesehen, dass der Teilnehmende seine Nutzer-Identifikationsnummer der entsprechenden Veranstaltung angibt, damit die Notizen von Server geladen werden können.

5 Richtlinienkonformität der Erweiterungen

Während der Erweiterung von BigBlueButton wurden regelmäßig Tests mit dem Screenreader Jaws durchgeführt, um dessen Unterstützung zu gewährleisten. Die implementierten Erweiterungen erhöhen die Zugänglichkeit von BigBlueButton, indem Richtlinien (vgl. Abschnitt 2.1) zur barrierefreien Gestaltung virtueller Klassenzimmer beachtet werden.

5.1 Erfüllte Richtlinien

Durch die Verbesserung der Screenreader-Kompatibilität in Bezug auf Tastaturunterstützung, Tabulator-Reihenfolge und Beschreibungen für Bedienelemente wird der Umfang wahrnehmbarer Inhalte erhöht (Richtlinie 1, Abschnitt 2.1) und die Navigation mittels Tastatur gewährleistet (Richtlinie 2). Die Konfigurationsmöglichkeit von Audio-Signalen, Layout und Tastaturkürzeln sowie von der Darstellung des Aktivitätsprotokolls ermöglicht eine Anpassung an persönliche Bedürfnisse (Richtlinie 8).

[7] http://popcornjs.org/



Abb. 4: Aufzeichnung einer Veranstaltung mit Aktivitätsprotokoll. Rote Hervorhebung: Kennzeichnung der Bereiche: 1 Folienübersicht, 2 Präsentationsbereich, 3 Reiter mit Chat, Notizen, Aktivitätsfenster und Einstellungen, 4 Abspielkontrolle

Das Aktivitätsprotokoll erfasst alle Ereignisse und textuellen Informationen und erlaubt die Veränderung der Schriftgröße für Teilnehmende mit Sehbeeinträchtigung und die Einschränkung des Informationsumfangs über Filter. Damit erleichtert es die Wahrnehmung der Informationen. Durch die Auflistung der Ereignisse in chronologischer Reihenfolge mit Zeitstempel wird die Orientierung (Richtlinie 3) unterstützt. Durch die Kommandozeile wird eine aktive Teilhabe an der Veranstaltung ermöglicht (Richtlinie 4) und die soziale Präsenz gefördert (Richtlinie 7), da u. a. Chat-Nachrichten verfasst, Audio- und Videoübertragung gesteuert und der Teilnehmenden-Status verändert bzw. abgefragt werden kann. Das Aktivitätsprotokoll und die Aufzeichnung unterstützen weiterhin das Verfolgen der Veranstaltung in eigener Geschwindigkeit (Richtlinie 6), da im Protokoll verpasste Inhalte nachvollzogen und bei dem Abspielen der Aufzeichnung bei Bedarf Pausen gemacht werden können.

Die kollaborative Beschreibungsmöglichkeit von Elementen verbessert die Wahrnehmbarkeit der präsentierten Objekte auf dem Whiteboard (Richtlinie 1). Da alle Teilnehmenden in der Lage sind, Beschreibungen hinzuzufügen und die von anderen zu bearbeiten, können Teilnehmende ohne Beeinträchtigung blinden Teilnehmenden Hilfestellung (Richtlinie 9) leisten, die Bedeutung gezeichneter Inhalte zu erfassen.

Durch das Notiz-Modul wird die Notwendigkeit eines Anwendungswechsels zum Notizenmachen verringert. Zukünftig soll es die Möglichkeit bieten, über den Zeitstempel, eine Verbindung zwischen der Notiz und einem Veranstaltungsinhalt herzustellen (Richtlinie 5). Das Laden von Notizen anderer Teilnehmenden in eine Veranstaltungs-Aufzeichnung (Richtlinie 9) kann das Verständnis der Veranstaltung fördern, wenn ein Teilnehmender selbst nicht in der Lage war, eine eigene Mitschrift anzufertigen.

5.2 Mögliche Erweiterungen

Neben der Erweiterung von BigBlueButton um weitere alternative Konzepte (vgl. Abschnitt 2.4), sind folgende Verbesserungen der neuen Funktionen wünschenswert:

- Erfassen von Alternativtexten von Folien-Abbildungen im Aktivitätsprotokoll
- Erweiterung des Notiz-Moduls um Zeitstempel, Listenansicht, Titelgenerierung und automatischer Speicherfunktion
- Integration der Notizen in das Aktivitätsprotokoll, so dass Notizen über die Kommandozeile erstellt und angezeigt werden können
- Verbesserung der Sicherheit beim Notizen-Zugriff, da die Nutzer-Identifikationsnummern leicht erraten werden können
- Lokalisierung der Beschreibungen für Bedienelemente und Audio-Signale für weitere Sprachen, um die Verständlichkeit des Screenreaders zu verbessern
- Erweiterung der verfügbaren Audio-Signale für Ereignisse
- Erweiterung der Aufzeichnung um das Beschreibungsmodul

6 Zusammenfassung und Ausblick

Die reine Möglichkeit der Bedienung eines virtuellen Klassenzimmers mit technischen Hilfsmitteln wie Screenreader und Braille-Zeile ist nur die Basis für dessen Verwendbarkeit. Erst die Erweiterung um weitere Unterstützungs- und Beschreibungsmechanismen führt zu Barrierefreiheit, wobei in einem Lehr-/Lernkontext auch immer die didaktische Aufbereitung Einfluss auf die gleichberechtigte Teilhabe hat (vgl. [SC10]).

Die prototypische Erweiterung des virtuellen Klassenzimmers BigBlueButton umfasst die Neuentwicklung des Notiz-, Aktivitätsprotokoll-, Beschreibungs- und Konfigurations-Moduls sowie die Verbesserung der Screenreader-Kompatibilität und Erweiterung des Funktionalität der Aufzeichnung (Abb. 5). Durch diese Erweiterungen wird die Zugänglichkeit für blinde, aber auch sehende, Teilnehmende, entschieden erhöht. Während der Entwicklung wurden kontinuierliche Tests bezüglich der Funktionalität und mit dem Screenreader Jaws durchgeführt. Die Bestätigung der Bedienbarkeit durch blinde Teilnehmende im Rahmen eines Benutzertests steht noch aus.

Aufgrund einer guten Dokumentation des Open-Source-Projekts lag zu Beginn die Schlussfolgerung einfacher Anpassungsmöglichkeiten nahe. Installation und Kompilation der Entwicklungsumgebung von BigBlueButton stellten sich jedoch als zeitaufwendig und fehleranfällig heraus, sodass mache Funktionalität aufgrund mangelnder Zeit noch nicht vollständig umgesetzt werden konnte. Nach der Vervollständigung der entwickelten Erweiterungen ist eine Übergabe an die BigBlueButton-Community geplant.

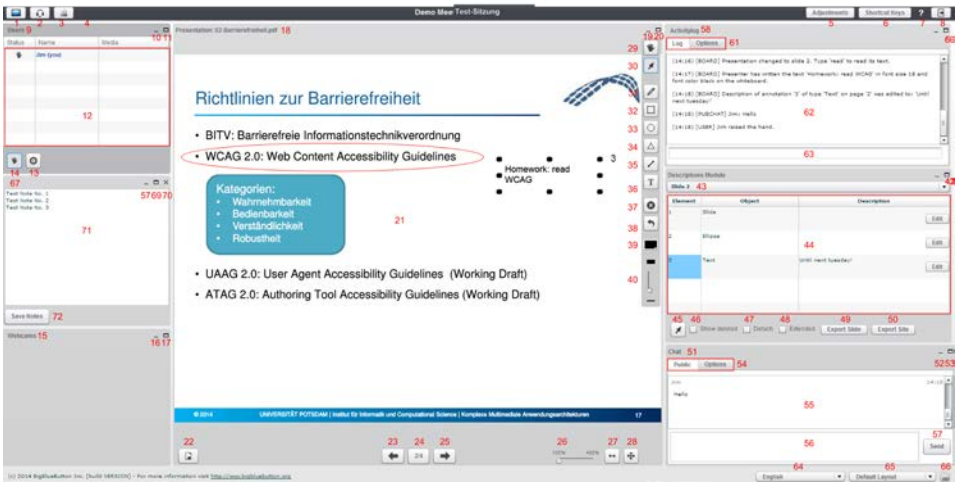


Abb. 5: Erweiterte Bedienoberfläche von BigBlueButton mit Tabulator-Reihenfolge in rot

Danksagung

Unser Dank gilt den weiteren Projektmitgliedern Robert Fruth, Manuel Hartmann, Marlene Karlapp, Mohammad Katani, Christian Klein, Jorge Elias Kuffatti Alvarado, Fritz Rose, Eckehart Synnatzschke, Volker Wölfert und Manuel Zedel, die durch ihre Mitarbeit diese Erweiterungen ermöglicht haben.

Literaturverzeichnis

[BBB13] BigBlueButton: 0.81Overview, https://code.google.com/p/bigbluebutton/wiki/081Overview#Accessibility, 2013, Stand 16.04.2015.

[BBB14] BigBlueButton: 090Accessibility. Implementation of accessibility in BigBlueButton 0.9.0-beta, https://code.google.com/p/bigbluebutton/wiki/090Accessibility, 2014, Stand: 16.04.2015.

[BBB15] BigBlueButton, http://www.bigbluebutton.org, Stand: 16.04.2015.

[Bl14] Blackboard Inc.: Blackboard collaborate web conferencing Version 12.6 – Accessibility Guide for Participants, 2014, http://www.blackboard.com/docs/documentation.htm?DocID=611012PDF, Stand 16.04.2015.

[DIN06] DIN EN ISO 9241-110:2006-03, Ergonomie der Mensch-System-Interaktion – Teil 110: Grundsätze der Dialoggestaltung (ISO 9241-110:2006).

[Fr10] Freire, A. et al.: Revealing the Whiteboard to Blind Students: An Inclusive Approach to Provide Mediation in Synchronous E-Learning Activities. Computers & Education 54/4, S. 866–876, 2010.

[FH12] Ferdinand, P.; Heckmann, P.: Plattformen. In (Haake, J. M.; Schwabe, G.; Wessner, M. Hrsg.): CSCL-Kompendium 2.0. Lehr- und Handbuch zum computerunterstützten kooperativen Lernen. Oldenbourg, München, S. 163–186, 2012.

[IMS04] IMS Global Learning Consortium: IMS Guidelines for Developing Accessible Learning Applications, Section 7, http://www.imsglobal.org/accessibility/accessiblevers/sec7.html, 2004, Stand: 16.04.2015.

[LGPL15] GNU Lesser General Public License, https://www.gnu.org/licenses/lgpl.html, Stand 16.04.2015.

[Kö14] Köhlmann, W.: Richtlinien zur barrierefreien Gestaltung virtueller Klassenzimmer für Blinde. i-com 13/3, S. 3–11, 2014.

[Kö15] Köhlmann, W.: Barrieren bei der Bedienung virtueller Klassenzimmer durch Blinde – Ergebnisse einer Analyse. Erscheint in: „Trendy, hip und cool“: Auf dem Weg zu einer innovativen Hochschule? Blickpunktreihe der DGHD, 2015.

[KL15] Köhlmann, W.; Lucke, U.: Alternative Concepts for Accessible Virtual Classrooms for Blind Users. In: Proc. 15th IEEE Int. Conf. on Advanced Learning Technologies (ICALT), Hualien, Taiwan 2015. IEEE, 2015 (im Druck).

[Me11] Mesiti, M. et al.: Collaborative Environments: Accessibility and Usability for Users with Special Needs. In (Pardede, E. Hrsg.): Community-Built Databases. Research and Development. Springer, Berlin, S. 319–340, 2011.

[Mo11] Mori, G. et al.: Making "Google Docs" User Interface More Accessible for Blind People: Proceedings of the First International Conference on Advances in New Technologies, Interactive Interfaces, and Communicability. Springer, Berlin, S. 20–29, 2011.

[SCM14] Santarosa, L.; Conforto, D.; Machado, R. P.: Whiteboard: Synchronism, Accessibility, Protagonism and Collective Authorship for Human Diversity on Web 2.0. Computers in Human Behavior 31, S. 591–601, 2014.

[SK15] Schäkel, C.; Köhlmann, W.: Programmatic Availability of Virtual Classrooms for Assistive Technologies. In: Proc. 8th Int. Conf. on Pervasive Technologies Related to Assisive Environments. 2015 (im Druck).

[SC10] Seale, J.; Cooper, M.: E-Learning and Accessibility: An Exploration of the Potential Role of Generic Pedagogical Tools. In Computers & Education, 54, S. 1107–1116, 2010.

[TC08] Talking Communities: Talking Communities V7.42 Voluntary Product Accessibility Template, http://talkingcommunities.com/vat/, 2008, Stand 16.04.2015.

[W3C08] World Wide Web Consortium: Web Content Accessibility Guidelines (WCAG) 2.0, http://www.w3.org/TR/2008/REC-WCAG20-20081211/, 2008, Stand 22.04.2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Meldestelle für digitale Barrieren

Christian Bühler[1], Christian Radek[2] und Arbeitsgruppe[3] der TU Dortmund

Abstract: Die Meldestelle für digitale Barrieren nimmt sich des Problems des barrierefreien Zugangs aus Nutzersicht – da, wo in der Praxis Schwierigkeiten auftreten – an. Eine solche verbraucherorientierte Komponente ist eine wichtige Ergänzung zu gesetzgeberischen Maßnahmen. Die Anbieter erhalten konkrete Rückmeldungen und können gezielt Schwächen ihrer Angebote verbessern. Im Rahmen einer Kooperation mit der Fakultät für Rehabilitationswissenschaft der TU Dortmund wurde eine Untersuchung zu den Themen Bekanntheit der Meldestelle und Motivation von Nutzern durchgeführt. Dazu wurden Expertinnen und Experten interviewt und aus den so gewonnenen Erkenntnissen ein Katalog mit Handlungsempfehlungen erstellt.

Keywords: Meldestelle, digitale Barrieren, Motivation, Barrierefreiheit, Universelles Design

1 Einleitung

In der UN-Behindertenrechtskonvention [UN15] ist Barrierefreiheit eines der Grundprinzipien - Artikel 9 widmet sich eigens diesem Thema. Somit wird Barrierefreiheit als grundlegende Voraussetzung für Inklusion und Teilhabe in der Gesellschaft angesehen. Dabei ist - neben dem Zugang zu Gebäuden und Verkehrsmitteln - der barrierefreie Zugang zu Medien, Information und Kommunikation ein für die Informations- und Wissensgesellschaft zunehmend wichtiger Bereich. Wer heute davon ausgeschlossen wird, dem bleibt die volle Teilhabe am Leben in unserer Gesellschaft verwehrt. Barrierefreiheit sorgt im Sinne eines universellen Designs [UD15a, UD15b] dafür, dass digitale Medien (Webangebote, Online-Dokumente, Software, Apps, Informationssysteme und Automaten) von einer größtmöglichen Gruppe von Menschen unabhängig und selbstbestimmt genutzt werden können. Nur vergleichsweise wenige Internetauftritte werden durch die aktuelle deutsche Gesetzgebung zur Barrierefreiheit direkt verpflichtet. Zum Beispiel ist die Barrierefreie-Informationstechnik-Verordnung (BITV) 2.0 [IT11] nur für die Behörden der Bundesverwaltung bindend. Auf Länderebene und im kommunalen Bereich greifen ähnliche Regelungen. Für die privaten Anbieter digitaler Medien gibt es keine derartige Verpflichtung, diese Medien barrierefrei anzubieten. Außerdem sieht die Barrierefreie-

[1] Technische Universität Dortmund, Emil-Figge-Straße 50, 44227 Dortmund, c.buehler@reha-technologie.de
[2] Bundesarbeitsgemeinschaft Selbsthilfe von Menschen mit Behinderung, chronischer Erkrankung und ihren Angehörigen e.V. (BAG SELBSTHILFE), Kirchfeldstraße 149, 40215 Düsseldorf, christian.radek@bag-selbsthilfe.de
[3] Lara Bergmann, Ramona Böhm, Rebecca Dettling, Lena Gianmoena, Nina Loddenkemper, Marcia Meier, Kathrin Olschewski, Jessica Pudysz, Ricarda Rüger, Romina Rüger, Lena Schulenberg und Deborah Wasem.

Informationstechnik-Verordnung (BITV) 2.0 keine wirkungsvollen Kontrollmechanismen oder gar Sanktionen bei Nichterfüllung der gesetzlichen Vorgaben vor. In diesem relativ rechtsfreien Raum (bezogen auf die Barrierefreiheit) vertritt die Meldestelle für digitale Barrieren die Interessen von Menschen mit Behinderungen, die bei der Nutzung digitaler Medien auf Barrieren stoßen. Gleichzeitig will sie den Anbietern helfen, bessere Informationsangebote für alle zur Verfügung zu stellen.

Die Meldestelle für digitale Barrieren [Bm15] wurde 2006 im Rahmen des „Aktionsbündnisses für barrierefreie Informationstechnik“ (AbI) (2002 - 2009) [AI15] als Meldestelle für Webbarrieren ins Leben gerufen. Im Rahmen des Projektes „Digital informiert – im Job integriert“ (Di-Ji) (2010 - 2014) [DJ15] wurde die Meldestelle auf das gesamte Spektrum digitaler Medien erweitert. Meldungen über Barrieren können telefonisch, via E-Mail oder über Webformulare bzw. eine Browser-Erweiterung (Firefox) gemeldet werden. Die Meldestelle setzt sich daraufhin mit dem Anbieter des digitalen Angebotes, in dem die Barriere auftritt, in Verbindung, bittet darum, die Barriere abzubauen und weist auf verfügbare Hilfs- und Beratungsangebote zur Umsetzung hin [BL15].

Es gibt weltweit ähnliche Angebote, die sich jedoch in Art und Umfang der Aktivitäten unterscheiden und zum Teil auf spezifische Zielgruppen fixiert sind (z.B. Fix the Web [FW15], Social Accessibility Project [SA15]). Diese sind auf die Gegebenheiten in den jeweiligen Ländern im Hinblick auf nationale Gesetze, Sprache und Kultur ausgerichtet und weisen unterschiedliche Überschneidungen mit der deutschen Meldestelle auf.

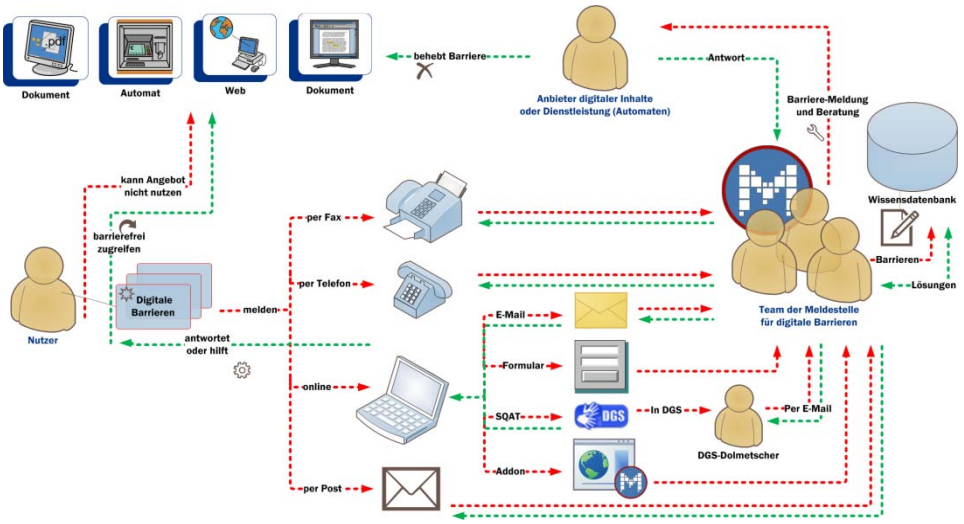


Abb 1: Abläufe bei der Bearbeitung von Barriere-Meldungen

Grundsätzlich lassen sich viele Barrieren in digitalen Medien leicht vermeiden. Wenn man zum Beispiel die Erstellung von Webseiten betrachtet, dann bietet HTML eine

Reihe von Funktionen an, die, wenn man sie so verwenden würde wie ursprünglich gedacht, Barrieren gar nicht erst entstehen lassen. Dazu zählen zum Beispiel Strukturelemente wie Überschriften. Überschriften dienen dazu, ein Dokument hierarchisch zu gliedern. Meldungen zeigen, dass es oft weder eine Überschrift der ersten Überschriftenebene noch irgendeine Art der Gliederung gibt.

Viele Meldungen betreffen Alternativtexte zu Abbildungen. In HTML lässt sich zu einem Bild ein Alternativtext angeben. Der Alternativtext wird genutzt, wenn das dazugehörige Bild nicht dargestellt werden kann. Das kann verschiedene Gründe haben: Server-Probleme, eine falsche Pfadangabe, Bilder sind im Browser deaktiviert oder eben eine "Lesehilfe" für Benutzer mit Sehschwäche oder Blindheit. Der Alternativtext ist eine Pflichtangabe und muss die Abbildung adäquat beschreiben. Allerdings gibt es eine Reihe von Ausnahmen. Dient die Abbildung nur der Dekoration, das heißt die Abbildung ist für das Verständnis der Webseite nicht wichtig, kann der Alternativtext entfallen.

Weitere Meldungen von Barrieren betreffen unter anderem unzureichende Farbkontraste, Linktexte, die außerhalb des Kontexts keinen Sinn ergeben (z. B. mehr …, weiter…), Festlegung auf ein Eingabegerät (Maus oder Tastatur) und audiovisuelle Medien ohne Untertitelung und ggf. Audiodeskription.

Eine oft gemeldete Barriere, die vor allem blinde und sehbehinderte aber auch alte Menschen von der Nutzung ausschließt, betrifft CAPTCHAs. CAPTCHAs bestehen meist aus verschnörkelten Buchstaben und Zahlenkombinationen, die erkannt und in ein Textfeld übertragen werden müssen. (Das Akronym CAPTCHA bedeutet „Completely Automated Public Turing test to tell Computers and Humans Apart", also ein komplett automatischer öffentlicher Turing-Test zur Unterscheidung von Computern und Menschen. CAPTCHAs werden verwendet, um Bestell- und Kontaktformulare gegen bösartige automatische Anfragen abzusichern.) Die hinter den CAPTCHAs stehende Überlegung ist die, dass nur Menschen die gestellten Aufgaben lösen können. Bei grafischen CAPTCHAs werden aber Menschen mit Sehbeeinträchtigungen ausgeschlossen, da sie diese nicht erkennen können. Die Wirksamkeit von CAPTCHAs ist überdies zweifelhaft. So hat z.B. Google für „Street View" einen Algorithmus entwickelt, um verzerrte Hausnummern auf Fotografien zu erkennen. Mit Hilfe dieses Algorithmus lassen sich CAPTCHAs automatisch mindestens genauso gut lösen wie durch Menschen [Go14]. Trotzdem wird weiterhin an CAPTCHAs festgehalten, obwohl diese viele Menschen ausschließen und obwohl es andere Möglichkeiten (serverseitige Filter, Spam-Fallen oder Zeitstempel) [Re15] gibt, Formulare abzusichern. Auch Audio-CAPTCHAs oder logische CAPTCHAs stellen keine echten Alternativen dar, da stets Menschen (z.B. gehörlose Menschen oder Menschen mit Lernbehinderung) durch das CAPTCHA von der Nutzung ausgeschlossen werden.

Die Prinzipien, die für barrierefreie Webseiten gelten, lassen sich auf Software und Online-Dokumente übertragen [EN14]. Software muss demnach auch über ausreichende Farbkontraste verfügen, gleichermaßen mit Maus und Tastatur bedienbar sein und

multimodale Ausgaben unterstützen. In Dokumenten müssen analog zu HTML Überschriften, Tabellen und Abbildungen ausgezeichnet werden, Farbkontraste ausreichend sein und Abbildungen über Alternativtexte verfügen.

Auch wenn Automaten, Informations- und Serviceterminals auf den ersten Blick nicht zur Gruppe der digitalen Medien dazuzugehören scheinen, verfügen sie doch über mittels Informationstechnik realisierte Benutzeroberflächen. Im Kontext von Barrierefreiheit ist dabei von Interesse, wie die Benutzerschnittstelle gestaltet ist. Es wäre wichtig, ganz im Sinne eines universellen Designs, multimodale Ein- und Ausgabesysteme bereitzustellen. Während Verkehrsbetriebe mittlerweile dazu übergehen, ihre Informationsstelen an den Haltestellen mit einer Sprachausgabe auszustatten, ist die Kommunikation bei der Bahn in der Regel monomodal. Im schlimmsten Fall steht dann auf der Anzeigetafel „Bitte auf Ansage achten".

Auch bei Automaten (Fahrkarten- oder Geldautomat) steht die Forderung nach multimodaler Interaktion im Vordergrund. Neben der grafischen Benutzeroberfläche muss es für blinde Menschen eine Sprachausgabe geben. Ein Touchscreen kann nicht von allen bedient werden. Besonders problematisch sind physische Ein- und Ausgabevorgänge (EC-Karte, Fahrkarte, Geldmünzen und -scheine). Diese stellen unter Umständen hohe Anforderungen an das zielgerichtete Greifen, eine entsprechende Greifkraft und dies unter Umständen in einem eng begrenzten Zeitfenster (Entnahme der EC-Karte und der Geldscheine aus einem Geldautomaten).

2 Einführung in die Problematik

Es liegt in der Natur einer Meldestelle, dass diese grundsätzlich nur aufgrund von Meldungen aktiv wird. Im Falle der Meldestelle für digitale Barrieren ist damit sichergestellt, dass Barrieren nicht flächendeckend, sondern tatsächlich an den Stellen, an denen sie Anwender behindern, abgebaut werden.

Die Anzahl der eingehenden Meldungen hängt dabei von verschiedenen Faktoren ab. Ein Aspekt ist dabei die Art und Weise, wie Meldungen erzeugt werden können. Je einfacher eine Barriere gemeldet werden kann, desto öfter wird davon Gebrauch gemacht. Aus diesem Grund wurde viel Energie darauf verwandt, insbesondere den Meldeprozess weiter zu vereinfachen. Um weitere Faktoren zu identifizieren und zu bewerten, entstand eine Zusammenarbeit mit der Fakultät Rehabilitationswissenschaft der TU Dortmund. Konkret stand die Frage im Raum, wie sich die Motivation, den Meldeprozess zu starten und damit die Anzahl der eingehenden Meldungen, weiter steigern lässt.

Der Begriff „Motivation" wird dabei als „Zusammenspiel von Faktoren, die in der Person liegen (Motive, Bedürfnisse, Interessen, Ziele), und Faktoren, die in der Umwelt liegen (Gelegenheiten, Anforderungen, Anreize)" [Br13] definiert. Dies ist eine der Grundannahmen aus der Motivationspsychologie. Fehlt einer dieser Faktoren, bleibt ein bestimmtes Verhalten aus. Der Mensch muss also in der Umwelt auf Anreize treffen, die

mit den jeweiligen individuellen Präferenzen korrelieren, damit ein bestimmtes Verhalten stattfindet [Br13].

3 Methodik

Um die Arbeitsweise der Meldestelle kennenzulernen, entwickelte eine Arbeitsgruppe der TU Dortmund in einem ersten Schritt Personas, also fiktive Personen, die aber mit ihren Vorlieben, Einschränkungen und sonstigen Attributen eine spezifische Personengruppe repräsentieren [Co03]. Das Ziel bestand darin, ein geeignetes Set von Personas mit unterschiedlichen (physischen, sensorischen und kognitiven) Einschränkungen zu erzeugen. In einem zweiten Schritt meldeten dann Mitglieder der Arbeitsgruppe aus der Perspektive dieser Personas Barrieren, auf die sie auf Webseiten, bei Software und bei Automaten aus dem Themenkomplex Ausbildung, Beruf und Arbeitswelt stießen. Im Rahmen dieser Meldungen wurde der Meldeprozess analysiert und Stärken und Schwächen der Meldestelle für digitale Barrieren identifiziert. Ausgehend von dieser ersten Analyse wurden Expertinnen und Experten aus den Bereichen Psychologie, Marketing, Kommunikation, Gamification, Selbsthilfe und barrierefreiem Webdesign nach ihrer Einschätzung und Handlungsoptionen befragt. Die Erkenntnisse aus der Meldephase und den Interviews wurden in einem dritten Schritt in einen Katalog mit Handlungsempfehlungen überführt.

4 Ergebnisse

In der ersten Phase der Untersuchung wurden praktische Erfahrungen mit der Erzeugung von Meldungen und dem Meldeprozess gemacht. Dabei wurde festgestellt, dass sowohl die Webseite, über die die Meldungen erzeugt werden, als auch der Meldeprozess Raum für Verbesserungen bietet. Dazu zählen unter anderem: die Anordnung der Objekte auf den Webseiten; die verwendete Sprache, die nicht als Alltagssprache, sondern eher als technische Sprache charakterisiert wurde; die Beschriftung der Menüs und der Eingabefelder; die Rückmeldungen, die es im Rahmen der Bearbeitung der Meldungen gibt bzw. die es nicht gibt. Außerdem wurde angeregt, den Meldevorgang, also die Art und Weise wie eine Meldung eingegeben wird, in Form einer Anleitung im Webauftritt unterzubringen. Die Meldestelle wurde eher als unattraktiv und der Anreiz, die Angebote der Meldestelle wahrzunehmen, als gering beurteilt.

Deshalb wurden mithilfe der Interviews nach Möglichkeiten gesucht, die Motivation zur Nutzung bzw. bei der Nutzung der Meldestelle zu steigern. Auf Basis der Interviews wurden neun Kategorien (Belohnung/Gratifikation, Bekanntmachung/Aufmerksamkeit, Transparenz, Bedienung, Design, Multiplikatoren, Kommunikationsweg umkehren, an Moral appellieren, Erweiterung der Nutzungsmöglichkeiten) identifiziert, die weiter in Unterkategorien heruntergebrochen wurden. Im Folgenden werden einige Ergebnisse aus den Interviews dargestellt, die am häufigsten genannt und als am wichtigsten betont

wurden.

Übereinstimmend loben die Experten die Idee der Meldestelle, die das Problem des barrierefreien Zugangs aus Nutzersicht – da wo in der Praxis Schwierigkeiten auftreten – angeht. Eine solche verbraucherorientierte Komponente wird als wichtige Ergänzung zu gesetzgeberischen Maßnahmen angesehen. Die Anbieter erhalten konkrete Rückmeldungen und können gezielt Schwächen ihrer Angebote verbessern.

Hinsichtlich der Umsetzung stand zunächst der Aspekt der Bekanntheit der Meldestelle für digitale Barrieren im Fokus. Durch die Befragung von Expertinnen und Experten aus verschiedenen Bereichen wurde deutlich, dass die Meldestelle selbst in Fachkreisen, die sich mit dem Thema Barrierefreiheit in digitalen Medien beschäftigen, nur sehr wenigen bekannt ist. Wenn keine persönliche bzw. fachliche Zusammenarbeit mit dem Personal der Meldestelle bestand, so war die Meldestelle nicht bekannt. Daraus wurde abgeleitet, dass als erstes an der öffentlichen Bekanntheit der Meldestelle gearbeitet werden muss. Ein möglicher Schritt wäre zum Beispiel die Nutzung sozialer Medien (z. B. Facebook, Twitter).

Das Thema Transparenz der Meldestelle und des Melde- bzw. des Bearbeitungsvorgangs bildet die nächsten Kategorie der Ergebnisse. Für Außenstehende wirkt das System der Meldestelle undurchsichtig. Was genau nach einer Meldung abläuft, wer von wem kontaktiert wird und was mit den Daten derjenigen, die die Barrieren melden, geschieht, kann nicht direkt erkannt werden. Wer die Meldestelle verantwortet, ist zwar auf der Internetseite der Meldestelle einsehbar, jedoch nicht präsent genug und schwierig zu erreichen. Auf der Internetseite der Meldestelle muss deutlicher werden, wer für die Bearbeitung der Meldungen verantwortlich ist und was mit eingehenden Meldungen geschieht.

Weiterhin stellt sich bei Betrachtung der Meldestelle die Frage, ob es bereits Erfolge vorzuweisen gibt. Das System und die Idee der Meldestelle für digitale Barrieren sind nach einstimmiger Meinung der befragten Expertinnen und Experten als positiv und gesellschaftlich relevant zu bewerten. Trotz dieser Einschätzung werden erfolgreich bearbeitete Meldungen bisher auch dann nicht öffentlich gemacht, wenn sie zur Verbesserung der Barrierefreiheit in digitalen Medien beigetragen haben. Diejenigen, die Barrieren gemeldet haben, erfahren so also bestenfalls individuell (beispielsweise via E-Mail), ob ihre Meldung Erfolg hatte. Dadurch kann ein weiteres wichtiges Potenzial, mit dem die Meldestelle nach Meinung der Expertinnen und Experten die Motivation ihrer Besucher erhöhen könnte, nicht genutzt werden. Es handelt sich dabei um die Verwendung eines Belohnungssystems und Emotionalisierung. Ein Belohnungssystem für Melderinnen und Melder, die regelmäßig konstruktive Meldungen absetzen, könnte auf unterschiedliche Weise gestaltet werden. In den Interviews ist jedoch immer wieder deutlich geworden, dass eine Veröffentlichung von erfolgreichen Meldungen die Nutzerinnen und Nutzer digitaler Medien auf einem sehr wirksamen moralischen/emotionalen Weg ansprechen würde.

Einen weiteren Punkt stellt die Beteiligung von Multiplikatoren und Betroffenen dar. Die

Meldestelle muss in ihrer zukünftigen Arbeit noch mehr Wert darauf legen, eine Zusammenarbeit mit zum einen selbst betroffenen Menschen aber auch zum anderen mit verschiedenen Einrichtungen und Institutionen herzustellen. Nur auf diesem Weg kann nach Meinung der befragten Personen das System der Meldestelle für digitale Barrieren optimiert werden.

5 Diskussion

Zunächst kann festgestellt werden, dass eine Einrichtung wie die Meldestelle für digitale Barrieren zur Umsetzung von Barrierefreiheit eine wichtige Rolle als verbraucherorientierte Komponente spielen kann. Sie hat auch ein Potenzial zur Unterstützung der Anbieter, in ihrem Bemühen möglichst gute digitale Informationen und Interaktionen anzubieten. Dies muss jedoch besser kommuniziert und weiterentwickelt werden.

In dieser Hinsicht lassen sich die Ergebnisse der Untersuchung grundsätzlich in zwei Bereiche einteilen. Auf der einen Seite stehen die Aspekte der Gebrauchstauglichkeit des Webauftritts. Hier lassen sich Verbesserungen vergleichsweise leicht umsetzen. Auf der anderen Seite stehen die Maßnahmen zur Steigerung der Bekanntheit und Motivation. Umsetzungen in diesem Bereich sind potentiell schwieriger, da oft juristische und andere Überlegungen in die Entscheidungen (Veröffentlichung von Webseitenanbietern, Erfassung und Speicherung personenbezogener Daten) mit einfließen müssen. In diesem Zusammenhang geht es nicht nur um die Motivation der potentiellen Melderinnen und Melder, sondern auch um eine Akzeptanz der Anbieter, die auf Probleme ihrer Angebote aufmerksam gemacht werden.

Eine interessante Frage ist weiter, ob es vertretbar ist, dass sich eine Initiative, die Barrieren abbauen will, barrierebehafteter Instrumente bedient: Ist es beispielsweise legitim, neben dem barrierefrei zugänglichen Webauftritt der Meldestelle eine eher unzugängliche Facebook-Seite zu betreiben oder wird man dadurch unglaubwürdig?

Schließlich steht noch die Frage im Raum, wie man Barrierefreiheit auch bei kommerziellen Anbietern erreicht, also was passieren muss, dass die Meldestelle für digitale Barrieren wieder überflüssig wird? Reicht dort eine „Selbstverpflichtung“, positive action (Auszeichnungen, Preise, usw.) oder brauchen wir harte gesetzliche Regelungen?

Literaturverzeichnis

[AI15] Aktionsbündnis für barrierefreie Informationstechnik (AbI), http://www.wob11.de/, Stand: 15.06.2015.

[BL15] BITV-Lotse, http://www.bitv-lotse.de/, Stand 15.06.2015.

[Bm15] Meldestelle für digitale Barrieren, http://barriere-melden.de/, Stand 15.06.2015.

[Br13] Brandstätter, V. et al.: Motivation und Emotion. Allgemeine Psychologie für Bachelor, Springer, Berlin, Heidelberg, 2013.

[Co03] Cooper Journal: http://www.cooper.com/journal/2003/08/the_origin_of_personas.html, Stand: 15.06.2015.

[DJ15] Digital informiert – im Job integriert (Di-Ji), http://www.di-ji.de/, Stand: 15.06.2015.

[EN14] EN 301 549 V1.1.1 (2014-02): Accessibility requirements suitable for public procurement of ICT products and services in Europe, http://www.etsi.org/deliver/etsi_en/301500_301599/301549/01.01.01_60/en_301549v010101p.pdf , Stand: 15.06.2015.

[FW15] Fix the Web, http://www.fixtheweb.net/, Stand 15.06.2015.

[Go14] Goodfellow, I. J. et al.: Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks, 2014, http://arxiv.org/pdf/1312.6082v4.pdf, Stand 15.06.2015.

[IT11] Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (Barrierefreie-Informationstechnik-Verordnung -- BITV 2.0), BITV 2.0. In: Bundesgesetzblatt: Bundesanzeiger Verlag, S. 1843–1859, 2011.

[Re15] Reins, F: Hintergrundinformationen - CAPTCHAs und Barrierefreiheit. In: Bundesministerium für Arbeit und Soziales (BMAS) (Hg.): BITV-Lotse einfach teilhaben. Unter Mitarbeit von Annika Nietzio, Frank Reins, Birgit Scheer und Rainer Wallbruch. Online verfügbar unter http://www.bitv-lotse.de/BL/DE/3_Hintergrundinformationen /3_5_CAPTCHAs_und_Barrierefreiheit/3_5_captchas_und_barrierefreiheit_node.html , Stand: 15.06.15.

[SA15] Social Accessibility Project, http://www.research.ibm.com/social/projects_sap.shtml, Stand: 15.06.2015, https://www.youtube.com/watch?v=Nkqhcb1foPI, Stand: 15.06.2015.

[UD15a] The Principles of Universal Design, The Centre for Universal Design, NC State University, http://www.ncsu.edu/ncsu/design/cud/about_ud/udprinciples.htm, Stand: 15.06.2015.

[UD15b] Die Prinzipien des Universellen Designs (Übersetzung), http://ftb-esv.de/uniprinc.html, Stand: 15.06.2015.

[UN15] Convention on the Rights of Persons with Disabilities, http://www.un.org/disabilities/convention/conventionfull.shtml, Stand: 15.06.2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Inklusion in Social Media: Die Perspektive der Betroffenen

Ricardo Buettner [1]

Abstract: Der vorliegende Beitrag beleuchtet aus der Perspektive der Technologieakzeptanzforschung (UTAUT2 [VTX12]) das Phänomen der (Nicht-)Nutzung von Social Media durch Sehbehinderte bzw. Blinde, indem speziell für diese Zielgruppe der bestehende Forschungsstand durch einen Usability-Experten mittels Tiefeninterview reflektiert wurde. Die Ergebnisse zeigen ein hohes Potential für die Wirtschaftsinformatik-/ IS-Forschung hinsichtlich der Erhöhung des Verständnisses der tatsächlichen Zugangs- und Nutzungsbarrieren von Social Media.

Keywords: Social Media, Inklusion, Technologieakzeptanz

1 Problemstellung

Die Social Media haben sich in der letzten Dekade zu einer der bedeutendsten Kommunikationstechnologien für die Gesellschaft entwickelt [Bu13b, Bu15b].[3] Obwohl die barrierefreie Gestaltung der Social Media für die soziale Inklusion von laut WHO-Schätzung [Wo11] circa 1 Milliarde behinderter Menschen weltweit essentiell ist [Sö09, Ta11], wurde diese noch sehr unzureichend erforscht und umgesetzt, woraus ein großer Handlungsbedarf resultiert [JX09, Bu13a]. Dieser Handlungsbedarf zeigt sich sowohl im Fehlen funktionsfähiger und akzeptierter Social Media-Praxisapplikationen für Behinderte als auch in einem eher in der Tiefe begrenzten Forschungsspektrum im Themenfeld Inklusion Behinderter in Social Media.

Forschungsseitig liegen vorwiegend Untersuchungen zu den Social Media Nutzungswünschen (bspw. [Br13]) und zur tatsächlichen Nutzungshäufigkeit (bspw. [Am09, Be08a, Fu12, WL11]) durch spezifische Behindertengruppen vor. Als Ergebnis dieser Untersuchungen lässt sich konstatieren, dass Behinderte die populären Social Media (insb. Facebook, Yahoo!, YouTube, Twitter und LinkedIn) uneingeschränkt nutzen möchten, genau diese populären Social Media jedoch sehr hohe Zugangs- und Nutzungsbarrieren aufweisen.

Diese Ergebnisse zu den Social Media Nutzungswünschen und tatsächlichen Nutzungshäufigkeiten sind für die Information Systems (IS) Forschung grundlegend. Zum gestaltungsorientierten Abbau der Zugangs- und Nutzungsbarrieren sind jedoch tiefergehende Forschungsergebnisse notwendig, die ein besseres Verständnis der konkreten Barrieren ermöglichen. Die hierzu vorhandenen Arbeiten zeigen jedoch nur, dass (a)

[1] FOM University of Applied Sciences, Institute of Management & Information Systems, ricardo.buettner@fom.de

[3] LinkedIn hatte im Juni 2015 364.000.000 Nutzer weltweit, Xing hatte im März 2015 8.800.000 Nutzer im Kernmarkt D-A-CH und Facebook erreichte im März 2015 ca. 1.441.000.000 Anwender. Zur Begriffsabgrenzung Social Media und (Online) Social Network (Sites) u.ä. vgl. stellvertretend [bE08b, Mi07].

die populären Social Media nicht behindertengerecht gestaltet sind [Be08a], (b) mobile Oberflächen aufgrund ihrer Komplexitätsreduktion bevorzugt werden [WL11], sowie (c) funktionale, soziale und individuelle Barrieren bestehen [DCP07]. Hier besteht deutlicher Forschungsbedarf für die Wirtschaftsinformatik, um das Verständnis der Zugangs- und Nutzungsbarrieren zu erhöhen.

Wie die Untersuchung von Douglas, Corcoran und Pavey [DCP07] gezeigt hat, bestehen die Zugangs- und Nutzungsbarrieren von Social Media für behinderte Menschen aus einem Konglomerat aus funktionalen, sozialen und individuellen Hindernissen bzw. Gegebenheiten. Das legt zum einen nahe, dass die vorwiegend funktional orientierten Ansätze spezifischer Social Media Schnittstellen für Behinderte (stellvertretend EasyChirp.com für Twitter als Praxisapplikation und [Sp12] als konzeptionelles Framework) zu kurz greifen. Zum anderen ist das in der Untersuchung von Douglas, Corcoran und Pavey [DCP07] gefundene Konglomerat aus funktionalen, sozialen und individuellen Hindernissen bzw. Gegebenheiten der Technologieakzeptanzforschung bekannt [BDW13], siehe insbesondere UTAUT [Ve03] und UTAUT2 [VTX12].

Der vorliegende Beitrag beleuchtet deshalb aus der Perspektive der Technologieakzeptanzforschung das Phänomen der (Nicht-)Nutzung von Social Media durch behinderte Menschen, indem speziell für die Zielgruppe Sehbehinderter bzw. Blinder der bestehende Forschungsstand durch einen Usability-Experten mittels Tiefeninterview reflektiert wird. Dieser Usability-Experte ist selbst blind und kann als Betroffener die Wirklichkeit sehbehinderter bzw. blinder Menschen besser wahrnehmen als nicht Sehbehinderte. In dieser Expertenperspektive liegt ein hohes Potential für die Wirtschaftsinformatik-/IS-Forschung, das Verständnis der tatsächlichen Zugangs- und Nutzungsbarrieren von Social Media für behinderte Menschen zu erhöhen.[4] Von den laut WHO-Schätzung [Wo11] circa 1 Milliarden behinderten Menschen weltweit sind ca. 285 Mio. Menschen sehbehindert, davon ca. 39 Mio. blind [Wo14]. Die Verbesserung der Zugangs- und Nutzungsbarrieren für Sehbehinderte bzw. Blinde hat demnach eine hohe gesellschaftliche Wirkung. Deshalb liegt der Fokus des Beitrags auf dieser Zielgruppe.

Der Beitrag ist wie folgt aufgebaut: Nach der Skizzierung der Problemstellung in diesem Abschnitt 1, zeigt der kommende Abschnitt 2 den aktuellen Forschungsstand zur ITK-Nutzung durch Behinderte. Darauf folgend werden im Abschnitt 3 die methodischen Überlegungen zur Untersuchung vorgestellt. Die Darstellung der Ergebnisse des Tiefeninterviews erfolgt in Abschnitt 4. Schließlich fasst Abschnitt 5 die Ergebnisse der Arbeit zusammen, kritisiert diese und gibt einen Ausblick auf weiteren Forschungsbedarf.

2 Aktueller Forschungsstand zur ITK-Nutzung durch Behinderte

Die hohen Potentiale der ITK-Nutzung für Behinderte werden bereits seit langer Zeit diskutiert. So legten beispielsweise Frolick, Wilkes und Urwiler [FWU93] vor 20 Jahren

[4] Da die Informatik/Wirtschaftsinformatik-Community die soziale Inklusion sehr aktiv fördert (u.a. W3C WAI Initiative) und hierzu bereits erfolgreich jährliche Konferenzen etabliert hat (Bsp. ICCHP, AAATE Conference, CSUN Conference oder auch der hiesige Workshop „Teilhabe an der allgegenwärtigen Kommunikation“) sollen die Ergebnisse auf dem Workshop „Teilhabe an der allgegenwärtigen Kommunikation“ vorgestellt werden.

bereits eine Untersuchung zur Nutzung von Tele-Arbeitsplätzen in der USA vor und wiesen auf die hohe Bedeutung dieser Arbeitsplätze für Behinderte hin. Ebenso wird regelmäßig die Notwendigkeit, legislativer und politischer Initiativen bzw. Regularien für diesen Themenbereich vorgebracht (u.a. [D'07, Ja06]). Zudem wird seit dem Entstehen des Internets auch die Web Usability und Barrierefreiheit in zahlreichen Arbeiten diskutiert (stellvertretend [HKJ08]).

Im Ergebnis dieser Bemühungen ist die ITK-Nutzung durch Behinderte in der Breite und Tiefe prinzipiell gut erforscht. Es liegen fast flächendeckend Arbeiten für definierte Anwendungsfelder und spezifische Behindertengruppen vor (stellvertretende Beispiele: für Sehbehinderte bzw. Blinde allgemein [WL09, WL11], speziell zur Nutzung von Lernsystemen durch Sehbehinderte und Blinde [RG90]; für geistig Behinderte [Li07]; allg. Literaturüberblick [Si07]). Zudem existieren verschiedene Hilfs- und Unterstützungsnetzwerke. Beispielsweise ermöglicht PatientsLikeMe.com das Auffinden von Personen mit ähnlichen Krankheitsbildern und den anschließenden Austausch mit diesen Patienten. Spezifische Forschungsarbeiten zur Social Media Nutzung durch Behinderte sind allerdings noch überschaubar.

2.1 Social Media Nutzung durch Behinderte

Prinzipiell lässt sich konstatieren, dass seitens der Social Media Unternehmen der Barrierefreiheit zu wenig Aufmerksamkeit geschenkt wurde. So analysierte die U.S. National Computing and Disability Charity „AbilityNet" bereits 2007 die sozialen Netzwerke Bebo, Facebook, MySpace, Yahoo! und YouTube und kam zu dem Ergebnis, dass keines dieser Netzwerke körperlich behinderten Menschen einen einigermaßen nutzerfreundlichen Zugang ermöglicht [Be08a]. Ein wesentliches Problem war beispielsweise durchgängig die Verwendung von Captchas, die als Social Media Barriere immer wieder kritisiert wird (u.a. [Go07, WL09]), jedoch größtenteils weiterhin alternativlos genutzt wird.

Neben diesen wenigen empirischen Arbeiten gibt es jedoch bereits designorientierte Ansätze: So zeigen Spiliotopoulos u.a. [Sp12] beispielsweise, wie eine Social-Media-Schnittstelle speziell für behinderte Menschen aussehen kann. Zudem existieren bereits erste Plattformen, die für spezifische Netzwerke Schnittstellen für behinderte Menschen anbieten (Bsp. EasyChirp.com für Twitter).

2.2 Social Media Nutzung speziell durch Sehbehinderte und Blinde

Die WHO schätzt, dass weltweit 285 Mio. Menschen sehbehindert sind, davon sind 39 Mio. blind [Wo14]. Diese große Zielgruppe wurde hinsichtlich der allgemeinen ITK-Nutzung prinzipiell gut untersucht (u.a. [RG90, WL09, WL11]). Im Ergebnis zeigt sich hierbei, dass ein wesentlicher Teil der Sehbehinderten und Blinden ITK nicht nutzt. Diejenigen Sehbehinderten, die jedoch Computer nutzen, verwenden diesen am häufigsten für Textverarbeitung (80%), E-Mails (73%), WWW Surfen (75%), Spiele/Freizeit/Photografie (53%), Tabellenkalkulation (41%) und Datenbanken (35%) [DCP07]. Seit dem

Aufkommen der Social Media und deren verstärkten Nutzung in der allgemeinen Bevölkerung (u.a. [bE08b, RRvB11, Bu14a, Bu14d, Bu15d]) steigt auch der Wunsch der Nutzung durch Sehbehinderte und Blinde.

So befragte 2009 die „American Foundation for the Blind“ [Am09] 62 Blinde zu ihren Nutzungshäufigkeiten und fand heraus, dass circa 50% Facebook, 33% Twitter, 25% LinkedIn und 25% MySpace nutzen. Fuglerud u.a. [Fu12] untersuchten 2010 in Norwegen das Nutzungsverhalten von Sehbehinderten (inkl. Blinder) mittels Telefoninterview (n=150) und Web Survey (n=69) mit dem Ziel, Nutzungsmotivation und -barrieren zu identifizieren. Die höchsten Nutzungsintensitäten erzielten Facebook (90%), Windows Messenger (80%), Skype (65%), Twitter (10%), YouTube (10%) und Google docs (50%). Obwohl es spezielle Netzwerke für Blinde gibt (u.a. InclusivePlanet.com), möchten die meisten Blinden die allgemeinen populären Social Media (insb. Facebook, Yahoo!, YouTube, Twitter und LinkedIn) wie die meisten anderen Menschen auch uneingeschränkt nutzen [Br13]. Der Fokus auf den populären Social Media lässt sich mit dem geringeren Sozialkapital kleinerer Netzwerke und der reduzierten Wahrscheinlichkeit von Kontakten erklären, u.a. [Br13, BML10, La12].

Es gibt demnach zwar empirische Untersuchungen zu den Social Media Nutzungshäufigkeiten von Blinden, es ist jedoch relativ wenig darüber bekannt, wie Blinde auf Social Media agieren, mit folgenden Ausnahmen: Brady u.a. [Br13] befragten 191 Blinde und führten Feldexperimente mit 23 Blinden durch. Es zeigte sich hierbei, dass Social Media behinderten Menschen helfen können, um an die Netzwerkpartner Fragen zu stellen. Beispielsweise können Blinde soziale Kontakte online bitten, ihnen Fotos zu beschreiben. [Br13] fand zudem heraus, dass Blinde den Wunsch nach reziproken Verhalten, Unabhängigkeit und nach Vermeiden von Hilflosigkeit bei der Nutzung von Social Media haben. Wentz und Lazar [WL11] verglichen die Nutzung von 15 Blinden auf der Facebook Website und der Facebook Mobile Site und konnten zeigen, dass die stationäre Website für Blinde deutlich umständlicher zu bedienen ist. In der Untersuchung des Social Media Nutzungsverhalten von Sehbehinderten (inkl. Blinde) durch Fuglerud u.a. [Fu12] war die Nutzungsrate mobiler Endgeräte ebenfalls durchweg sehr hoch (99% durchschnittlich). Die Reduktion der funktionalen Komplexität scheint ein Indikator für die Beliebtheit mobiler Anwendungen zu sein. Douglas, Corcoran und Pavey [DCP07] befragten basierend auf der WHO-Klassifikation [Wo01] in Großbritannien 960 Sehbehinderte zur deren Computernutzung mit dem Ziel, potentielle Barrieren der sozialen Inklusion zu identifizieren. Im Ergebnis kristallisierten sich (a) fehlende funktionale Relevanz, (b) soziale Barrieren und (c) individuelle Faktoren (insbesondere die Sehbehinderung selbst) als wesentliche Barrieren der Computernutzung heraus.

Neben diesen empirischen Publikationen existieren gestaltungsorientierte Arbeiten (Bsp. Guidelines für die barrierefreie Gestaltung von Online Communities [JX09] und Social Media [JBS12]).

3 Methodische Überlegungen

Durchgängig weisen Studien (u.a. [Ca12, WL09]) darauf hin, dass die Gestaltung von barrierefreien Informationssystemen nur möglich ist, wenn die spezifischen Bedürfnisse von behinderten Menschen dem Gestalter bekannt sind. Wie der oben ausgeführte Forschungsstand gezeigt hat, bestehen die Zugangs- und Nutzungsbarrieren von Social Media für behinderte Menschen aus einem Konglomerat aus funktionalen, sozialen und individuellen Hindernissen bzw. Gegebenheiten. Diese Unterteilung ist der Technologieakzeptanzforschung bekannt (siehe insbesondere UTAUT [Ve03] und UTAUT2 [VTX12]).

Um das Verständnis der Zugangs- und Nutzungsbarrieren von Social Media für Sehbehinderte zu erhöhen, werden in Anlehnung an das Technologieakzeptanzmodell UTAUT2 [VTX12] folgende fünf Haupteinflussbereiche mittels Tiefeninterview eines selbst betroffenen Usability-Experten abgefragt: [1] Welche Zugangsbarrieren zu Social Media bestehen für Sehbehinderte?, [2] Wie beurteilen Sie die Nutzerfreundlichkeit von Social Media für Sehbehinderte? (vgl. 'effort expectancy'), [3] Welche Social Media Funktionen sind für Sehbehinderte besonders interessant? (vgl. 'performance expectancy'), [4] Wie wird die Nutzung von Social Media durch das soziale Umfeld unterstützt? (vgl. 'social influence/facilitating conditions'), [5] Wie viel Spaß haben Sehbehinderte bei der Nutzung von Social Media? (vgl. 'hedonic motivation') [VTX12, S. 160, Abb. 1].

Das Tiefeninterview folgt den Empfehlungen von Bortz und Döring [BD09].

4 Relevante Auszüge aus dem Tiefeninterview des Betroffenen

4.1 Charakterisierung des Betroffenen und seiner Erfahrungsperspektive

Der interviewte Usability-Experte arbeitete nach Ausbildung und Studium fünf Jahre bei einer Kölner Onlinemarketing-Firma, die auf die Umsetzung von barrierefreien Web-Portalen spezialisiert ist. Hier war er an der Umsetzung und Kontrolle von Web-Portalen, CMS-Systemen und normalen Webseiten im Bezug auf Barrierefreiheit und Nutzerfreundlichkeit verantwortlich. Zum Zeitpunkt des Interviews war er selbstständiger Softwareentwickler und Webseitengestalter mit Schwerpunkt Barrierefreiheit und Nutzerfreundlichkeit für Medizin-, Hilfsmittel- und Softwareunternehmen. Der Usability-Experte erblindete mit dem fünften Lebensjahr zu 100%.

4.2 Zugangsbarrieren zu Social Media

„Die Zugangsbarrieren für Blinde und Sehbehinderte zu Social Media-Angeboten sind meist recht hoch. Eine erste Hürde ist oftmals schon die Registration. Bestimmte Links, Auswahlmenüs und Buttons sind nicht eindeutig beschriftet oder sogar gar nicht benannt. Oftmals werden zwar grafische Buttons verwendet, aber es wird vergessen, diese mit

beispielsweise Labels zu versehen. So kann man mit einem Screenreader sehr schnell an die Grenzen stoßen und ist beim Registrierungsprozess auf sehende Hilfe angewiesen."

„Beim Anmelden stößt man oft auf die nächste Hürde. Hier werden oft Sicherheitscodes abgefragt, die teilweise zwar auch als Audioformat angeboten werden, in der Regel aber nur schwer oder gar nicht verständlich sind. Mein eigener Versuch, mich bei LinkedIn zu registrieren, konnte auch nur mit sehender Hilfe durchgeführt werden. Nachdem ich dann mehrfach versucht hatte, mich anzumelden und immer an dem Sicherheitscode scheiterte, (Audio war für mich nicht verständlich) habe ich es bei LinkedIn aufgegeben."

„Viele Social Media-Angebote werden schnell weiterentwickelt und meiner Erfahrung nach spielt bei der Weiterentwicklung die Frage der Barrierefreiheit keine führende Rolle. Standardfunktionen wie das Hochladen eines Bildes ist dann schon für einen Blinden schwer möglich, da der Button 'hochladen' nicht mit einem Label versehen ist und da für einen Blinden nur 'button' steht. Das könnte aber auch ein Button für 'Verwerfen', 'Löschen' etc. sein."

4.3 Nutzerfreundlichkeit von Social Media

„Social Media sind aus Sicht von Blinden und Sehbehinderten nicht nutzerfreundlich. Das liegt meiner Meinung nach an den folgenden drei Hauptmerkmalen."

„1. Die Struktur: Oftmals ist die Struktur einer Webseite aus Sicht eines Screenreadernutzers nicht leicht verständlich. Als Sehender hat man sofort die Möglichkeit, sich einen Überblick über den Inhalt des Monitors zu machen. Wo ist die Navigation, wo ist die Suche, der Loginbereich etc. Als Blinder liest man den HTML-Code von oben bis unten durch. Hier fehlen oftmals so genannte 'Skiplinks', mit denen es möglich wäre, seitenintern beispielsweise direkt zum 'Loginbereich' zu springen. Für sehende Menschen spielt es keine sehr große Rolle wo ein Teil einer Webseite im Quellcode definiert wird. So könnte die Sitebar beispielsweise zum Beginn des HTML-Codes definiert werden und via CSS an die rechte Seite gesetzt werden. Für einen Screenreader ist es jedoch so, dass er ähnlich wie ein Suchmaschinen-Robot den Quellcode am Anfang beginnt und einfach durchliest. Somit wäre die im Beispiel genannte Sitebar zu Beginn der Seite für einen Blinden. Wenn es jetzt auch die so genannten Skiplinks nicht gibt, ist eine zügige Orientierung auf einer sich ständig ändernden Seite nahezu aussichtslos."

„2. Die Webtechniken: Viele der Webtechniken, die im Social Media-Umfeld verwendet werden, sind neu und werden daher ab und an von Browsern noch nicht regelkonform verwendet. Hier haben dann oft die Screenreader das Problem, dass sie gar nicht wissen können, wie damit umzugehen ist. [...] Allerdings habe ich auch die Erfahrung gemacht, dass die Webtechniken unterstützt würden, aber die Implementierung nicht sauber ist. So sind viele Javascript libs durchaus einsetzbar, wenn etwas mehr Augenmerk auf die Nutzerfreundlichkeit/Barrierefreiheit gelegt worden wäre. Hier werden sehr häufig auch Mythen weiterverbreitet, dass Javascript von Screenreadern gar nicht verstanden würde und somit alle Javascript-Lösungen für Blinde und Sehbehinderte nicht zu nutzen seien.

Diese Informationen sind jetzt falsch und waren höchstens zu Beginn der Entwicklung richtig.“

„3. Die Hilfetexte: Hilfestellungen in Form von Hilfetexten werden oftmals durch Webtechniken umgesetzt, die ausgerechnet nicht von Screenreadernutzern gelesen werden können. Somit fallen viele dieser Informationen ganz weg. Auch Dokumentation oder FAQs, die man ggf. auch als strukturiertes DAISY (navigierbares mp3) anbieten könnte, gibt es so gut wie nie.“

4.4 Zur Nutzung spezifischer Social Media Funktionalitäten

„Speziell für Blinde oder Sehbehinderte entwickelte Einsatzgebiete für Social Media könnten das Erkennen von Gegenständen sein. So gibt es den Versuch einer kanadischen Universität, eine Mustererkennung via App anzubieten, die von Menschen (via Social Media) unterstützt wird. Ziel ist, dass ein Blinder ein Photo mit seinem Smart Phone von einem Gegenstand oder einer beispielsweise Speisekarte machen kann und dann entweder von einer Mustererkennung (Software) oder einem Menschen ein Feedback darüber bekommt, was auf dem Photo zu sehen ist.“

„Mir ist bekannt, dass manche Blinde mit einer App für Facebook versuchen, ihre Orientierung sicherzustellen. Ein Beispiel: Ich weiß nicht genau wo ich bin, das erste mal in einer neuen Stadt und ich bin unsicher wo ich gerade bin. Die herkömmlichen Navigationsprogramme können mir nicht ganz genau die Hausnummer oder das Namensschild sagen. Nun kann ich ein Bild mit meinem Smartphone machen (Gebäude, Schild). Dieses Bild laden die blinden Nutzer dann zu Facebook hoch und ihre Freunde können ihnen genau sagen, was zu sehen ist. [...] Auch bestimmte Onlinespiele, die aus Social Media heraus gefüttert werden, könnten sinnvoll sein, da sie oftmals Geo-Daten enthalten. Diese sind zur Orientierung auch für Blinde und Sehbehinderte oftmals sehr hilfreich.“

„Allerdings wären Social Media-Angebote auch im Bezug auf eigenverantwortliches Arbeiten oder zur Arbeitssuche sehr hilfreich. Computer sind durchaus für viele behinderte Menschen zugänglich und bieten ein weites Spektrum an Berufschancen für die Betroffenen. Ein Ausschluss von Social Media ist somit eine Verminderung der Berufschancen, da viele Kontakte nicht entstehen oder nicht richtig gepflegt werden können.“

4.5 Unterstützung der Social Media Nutzung durch das soziale Umfeld

„Dem sozialen Umfeld ist meist gar nicht bewusst, dass behinderte Menschen das Internet nutzen können. Falls es doch bewusst ist, dann glauben viele Menschen, es sei eine Art Parallel-Netz. Dass behinderte Menschen auf den bekannten Webseiten und in den bekannten Social Media-Angeboten unterwegs sind, ist überhaupt nicht in der Mitte der Gesellschaft angekommen. Dass das soziale Umfeld von Behinderten deren Social Media Nutzung unterbindet, würde ich nicht sagen, aber unterstützt wird sie nicht. Sie ist meist gar kein Thema.“

4.6 Spaß bei der Nutzung

„Die Nutzung könnte besonderen Spaß machen. Für viele Menschen sind Social Media eine sehr große Chance, sowohl privat in Kontakt zu anderen Menschen zu kommen, als auch beruflich. Sowohl Privatleben als auch die Arbeit können Spaß machen, aber momentan sind die Social Media noch nicht soweit, dass sie wirklich Spaß machen. Der Stand heute ist eher, dass man kämpfen muss, um an die Inhalte zu kommen und zu partizipieren. Da Social Media eine immer größere Bedeutung erlangen, ist das wirklich problematisch."

5 Zusammenfassung, Kritik und weiterer Forschungsbedarf

Der vorliegende Beitrag beleuchtete aus der Perspektive der Technologieakzeptanzforschung (UTAUT2 [VTX12]) das Phänomen der (Nicht-)Nutzung von Social Media durch Sehbehinderte bzw. Blinde, indem speziell für diese Zielgruppe der bestehende Forschungsstand durch einen Usability-Experten mittels Tiefeninterview reflektiert wurde. Die Ergebnisse des Tiefeninterviews lassen sich strukturell wie folgt zusammenfassen: Hinsichtlich der Social Media Zugangsbarrieren wurde die Verwendung von schwer lesbaren – und im Falle eines Audioangebots schwer hörbaren – Sicherheitscodes(Captchas) sowie eine hohe Änderungsdynamik erkannt. Die schlechte Social Media Nutzerfreundlichkeit wird durch eine fehlende Navigationsstruktur, einer hohen Änderungsdynamik i.V.m. einer fehlenden Orientierung an Web-Standards zur Barrierefreiheit verursacht. Besonders interessante Social Media Funktionen betreffen die Möglichkeit, Fragen an andere Netzwerkpartner zu stellen (vgl. auch [Br13]), lokationsbasierte Dienste zu nutzen, sowie Social Media zur Arbeitssuche (ggf. i.V.m. eLancing/Crowdsourcing, siehe stellvertretend [AL13, ZZ14, Bu15a, Bu14c, Bu15d]) zu verwenden. Dem sozialen Umfeld Behinderter sind diese Möglichkeiten meist nicht bewusst, aufgrund der hohen Zugangs- und Nutzungsbarrieren ist der Spaßfaktor der derzeitigen Nutzung noch gering.

Kritisiert werden muss, dass lediglich ein Usability-Experte interviewt wurde. Da dieser Usability-Experte jedoch selbst blind ist und als Betroffener im Vergleich zu nicht Seh-behinderten die Wirklichkeit sehbehinderter bzw. blinder Menschen besser wahrnehmen kann, liegt in dieser Expertenperspektive ein hohes Potential für die Wirtschaftsinformatik-/IS-Forschung hinsichtlich der Erhöhung des Verständnisses der tatsächlichen Zugangs- und Nutzungsbarrieren von Social Media. Im nächsten Schritt sollen die vorerst identifizierten Zugangs- und Nutzungsbarrieren durch weitere Experten qualitativ ergänzt werden. Anschließend sollen quantitative Befragungen zum subjektiven Erleben durchgeführt werden. Diese Ergebnisse werden mittels quantitativer Verhaltensbeobachtung Sehbehinderter via Eye-Tracking [EMB12, Bu13c, Ec13, Bu14b, Bu15e, Bu15c] trianguliert und hinsichtlich der Theoriebildung ausgewertet.

Danksagung

Der Autor möchte den anonymen Gutachtern der INFORMATIK 2015 Konferenz für die wertvollen Hinweise herzlich danken. Die vorliegende Arbeit wurde teilweise durch das Bundesministerium für Bildung und Forschung (BMBF) unter dem Kennzeichen 03FH055PX2 gefördert.

In Erinnerung an CD.

Literaturverzeichnis

[AL13] Aguinis, Herman; Lawal, Sola O.: eLancing: A review and research agenda for bridging the science-practice gap. Human Resource Management Review, 23(1):6–17, 2013.

[Am09] American Foundation for the Blind: Survey Results on Social Networking. AccessWorld, 10(3), 2009.

[BD09] Bortz, Jürgen; Döring, Nicola: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Springer Medizin Verlag, Heidelberg, 4. edition, 2009.

[BDW13] Buettner, Ricardo; Daxenberger, Barbara; Woesle, Christian: User acceptance in different electronic negotiation systems - a comparative approach. In: In ICEBE 2013: Proceedings of the 10th IEEE International Conference on e-Business Engineering, September 11 - 13, Coventry, UK. IEEE CS Press, pp. 1–8, 2013.

[Be08a] Bearne, Suzanne: Social networks exclude disabled users through inaccessible sign-up. New Media Age, p. 11, 2008.

[bE08b] boyd, danah. m.; Ellison, Nicole B.: Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication, 13(1):210–230, 2008.

[BML10] Burke, Moira; Marlow, Cameron; Lento, Thomas: Social Network Activity and Social Well-Being. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '10, pp. 1909–1912, 2010.

[Br13] Brady, Erin L.; Zhong, Yu; Morris, Meredith Ringel; Bigham, Jeffrey P.: Investigating the Appropriateness of Social Network Question Asking as a Resource for Blind Users. In: CSCW '13 Proc. CSCW '13, pp. 1225–1236, 2013.

[Bu13a] Buettner, Ricardo: Social inclusion in eParticipation and eGovernment solutions: A systematic laboratory-experimental approach using objective psychophysiological measures. In: EGOV/ePart 2013: Proceedings of the Joint Conference of IFIP EGOV 2013 & IFIP ePart 2013, September 16-19, Koblenz, Germany, 2013. volume P-221 of Lecture Notes in Informatics (LNI). Gesellschaft für Informatik (GI), pp. 260–261, 2013.

[Bu13b] Buettner, Ricardo: The Utilization of Twitter in Lectures. In: INFORMATIK 2013: 43. Jahrestagung der Gesellschaft für Informatik, September 16-20, 2013, Koblenz, Germany. volume P-220 of Lecture Notes in Informatics (LNI). Gesellschaft für Informatik (GI), pp. 244–254, 2013.

[Bu13c] Buettner, Ricardo; Daxenberger, Barbara; Eckhardt, Andreas; Maier, Christian: Cognitive Workload Induced by Information Systems: Introducing an Objective Way of Measuring based on Pupillary Diameter Responses. In: Pre-ICIS HCI/MIS 2013 Proc. 2013. Paper 20.

[Bu14a] Buettner, Ricardo: A Framework for Recommender Systems in Online Social Network Recruiting. In: HICSS-47 Proc. pp. 1415–1424, 2014.

[Bu14b] Buettner, Ricardo: Analyzing Mental Workload States on the Basis of the Pupillary Hippus. In: NeuroIS '14 Proc. p. 52, 2014.

[Bu14c] Buettner, Ricardo: Crowdsourcing of a Human Resource Management Perspective: State of the Art, Challenges & Future Need for Research. Presentation at VHB '14 Conference, Leipzig, Germany, June 11-13, unpublished, 2014.

[Bu14d] Buettner, Ricardo: efficientRecruiting 2.0: Effizientes Recruiting von Fachkräften im Web 2.0. In: KnowTech 2014 Proc. BITKOM, pp. 119–127, 2014.

[Bu15a] Buettner, Ricardo: A Systematic Literature Review of Crowdsourcing Research from a Human Resource Management Perspective. In: HICSS-48 Proc. pp. 4609–4618, 2015.

[Bu15b] Buettner, Ricardo: Analyzing the Problem of Employee Internal Social Network Site Avoidance: Are Users Resistant due to their Privacy Concerns? In: HICSS-48 Proc. pp. 1819–1828, 2015.

[Bu15c] Buettner, Ricardo: Investigation of the Relationship Between Visual Website Complexity and Users' Mental Workload: A NeuroIS Perspective. In: Information Systems and Neuro Science: Gmunden Retreat on NeuroIS 2015. volume 10 of LNISO, pp. 123–128, 2015.

[Bu15d] Buettner, Ricardo: Personality as a predictor of business social media usage: An empirical investigation of XING usage patterns. Presentation at VHB '15 Conference, Vienna, Austria, May 27-29, unpublished, 2015.

[Bu15e] Buettner, Ricardo; Sauer, Sebastian; Maier, Christian; Eckhardt, Andreas: Towards ex ante Prediction of User Performance: A novel NeuroIS Methodology based on Real-Time Measurement of Mental Effort. In: HICSS-48 Proc. pp. 533–542, 2015.

[Ca12] Carrière, Véronique: ICT and cooperation in learning / teaching in visually handicapped situation. Procedia - Social and Behavioral Sciences, 46:701–705, 2012.

[D'07] D'Aubin, April: Working for Barrier Removal in the ICT Area: Creating a More Accessible and Inclusive Canada. Information Society, 23(3):193–201, 2007.

[DCP07] Douglas, Graeme; Corcoran, Christine; Pavey, Sue: The role of the WHO ICF as a framework to interpret barriers and to inclusion: visually impaired peoples views and experiences of personal computers. British Journal of Visual Impairment, 25(1):32–50, 2007.

[Ec13] Eckhardt, Andreas; Maier, Christian; Hsieh, J.J. Po-An; Chuk, Tim; Chan, Antoni B.; Hsiao, Antoni B.; Buettner, Ricardo: Objective measures of IS usage behavior under conditions of experience and pressure using eye fixation data. In: ICIS 2013 Proceedings: 34rd International Conference on Information Systems, December 15-18, 2013, Milan, Italy. 2013.

[EMB12] Eckhardt, Andreas; Maier, Christian; Buettner, Ricardo: The Influence of Pressure to Perform and Experience on Changing Perceptions and User Performance: A Multi-Method Experimental Analysis. In: ICIS 2012 Proc. 2012.

[Fu12] Fuglerud, Kristin Skeide; Tjøstheim, Ingvar; Gunnarsson, Birkir Rúnar; Tollefsen, Morten: Use of Social Media by People with Visual Impairments: Usage Levels, Attitudes and Barriers. In (Miesenberger, Klaus; Karshmer, Arthur; Penaz, Petr; Zagler, Wolfgang, eds): Computers Helping People with Special Needs, volume 7382 of Lecture Notes in Computer Science, pp. 565–572. Springer Berlin Heidelberg, 2012.

[FWU93] Frolick, Mark N.; Wilkes, Ronald B.; Urwiler, Robert: Telecommuting as a workplace alternative: an identification of significant factors in American firms' determination of work-at-home policies. Journal of Strategic Information Systems, 2(3):206–220, 1993.

[Go07] Goldie, Luan: Disabled users are denied access to Facebook, Google and Yahoo! New Media Age, p. 3, 2007.

[HKJ08] Hong, Soongoo; Katerattanakul, Pairin; Joo, Seok Jeong: Evaluating Government Website Accessibility: A Comparative Study. International Journal of Information Technology & Decision Making, 7(3):491–515, 2008.

[Ja06] Jaeger, Paul T.: Telecommunications policy and individuals with disabilities: Issues of accessibility and social inclusion in the policy and research agenda. Telecommunications Policy, 30(2):112–124, 2006.

[JBS12] Jaeger, Paul T.; Bertot, John Carlo; Shilton, Katie: Access Perspectives and Design Values in Government Social Media Usage. In: Proceedings of the 13th Annual International Conference on Digital Government Research. dg.o '12, pp. 216–222, 2012.

[JX09] Jaeger, Paul T.; Xie, Bo: Developing Online Community Accessibility Guidelines for Persons With Disabilities and Older Adults. Journal of Disability Policy Studies, 20(1):55–63, 2009.

[La12] Lampe, Cliff; Vitak, Jessica; Gray, Rebecca; Ellison, Nicole: Perceptions of Facebook's Value as an Information Source. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '12, pp. 3195–3204, 2012.

[Li07] Li-Tsang, Cecilia W. P.; Lee, Maggie Y. F.; Yeung, Susanna S. S.; Siu, Andrew M. H.; Lam, C. S.: A 6-month follow-up of the effects of an information and communication technology (ICT) training programme on people with intellectual disabilities. Research in Developmental Disabilities, 28(6):559–566, 2007.

[Mi07] Mislove, Alan; Marcon, Massimiliano; Gummadi, Krishna P.; Druschel, Peter; Bhattacharjee, Bobby: Measurement and analysis of online social networks. In: IMC '07 Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. ACM, pp. 29–42, 2007.

[RG90] Roberts, Andrew; Gough, Tom G.: Teaching information systems to the visually handicapped - a case history. Journal of Information Technology, 5(3):161–167, 1990.

[RRvB11] Richter, Daniel; Riemer, Kai; vom Brocke, Jan: Internet Social Networking: Research State of the Art and Implications for Enterprise 2.0. Business & Information Systems Engineering, 3(2):89–101, 2011.

[Si07] Sinclair, Stephen; Bramley, Glen; Dobbie, Louise; Gillespie, Morag: Social Inclusion and Communications: A Review of the Literature. Technical report, Center for Research into Socially Inclusive Services, 2007.

[Sö09] Söderström, Sylvia: Offline social ties and online use of computers: A study of disabled youth and their use of ICT advances. New Media & Society, 11(5):709–727, 2009.

[Sp12] Spiliotopoulos, Dimitris; Tzoannos, Efstratios; Stavropoulou, Pepi; Kouroupetroglou, Georgios; Pino, Alexandros: Designing User Interfaces for Social Media Driven Digital Preservation and Information Retrieval. In: Proceedings of the 13th International Conference on Computers Helping People with Special Needs. volume 1 of ICCHP'12, pp. 581–584, 2012.

[Ta11] Taylor, Anne: Social Media as a Tool for Inclusion. Technical report, Stiles Associates Inc., 2011.

[Ve03] Venkatesh, Viswanath; Morris, Michael G.; Davis, Gordon B.; Davis, Fred D.: User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly, 27(3):425–478, 2003.

[VTX12] Venkatesh, Viswanath; Thong, James Y. L.; Xu, Xin: Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly, 36(1):157–178, 2012.

[WL09] Wentz, Brian; Lazar, Jonathan: Email Accessibility and Social Networking. In (Ozok, A. Ant; Zaphiris, Panayiotis, eds): Online Communities and Social Computing, volume 5621 of Lecture Notes in Computer Science, pp. 134–140. Springer Berlin Heidelberg, 2009.

[WL11] Wentz, Brian; Lazar, Jonathan: Are Separate Interfaces Inherently Unequal? An Evaluation with Blind Users of the Usability of Two Interfaces for a Social Networking Platform. In: Proceedings of the 2011 iConference. iConference '11, pp. 91–97, 2011.

[Wo01] World Health Organization: International Classification of Functioning Disability and Health. Report, 2001.

[Wo11] World Health Organization & World Bank: World Report on Disability. Report, 2011.

[Wo14] World Health Organization: Visual impairment and blindness. Report, August 2014. Fact Sheet N282.

[ZZ14] Zhao, Yuxiang; Zhu, Qinghua: Evaluation on crowdsourcing research: Current status and future direction. Information Systems Frontiers, 16(3):417–434, 2014.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Development towards a generally applicable process to inspect and verify accessibility of web pages

Mathias Haimerl[1]

Abstract: Due to the onward process of inclusion of people with disabilities, a growing number of handicapped people start using the internet as a source of information. The World Wide Web Consortium (W3C) group, whose task it is to focus on web standards to make the web more accessible, named Web Accessibility Initiative (WAI), built a set of guidelines to achieve this objective. The Web Content Accessibility Guidelines (WCAG) 2.0[2], which is the current version of the ruleset, do unfortunately not provide enough techniques to cover most peoples' needs. Neither does it supply a practical method to test the compliance of its specifications. This paper provides a critical view on all guidelines of the WCAG, expands them to include more kinds of disabilities and recapitulates the revised rules with a set of practically applicable approaches to enable testing, using freeware or open source tools. The Examination of the WCAG is accomplished by investigating selected rules in terms of psychology as well as medical and educational science. Furthermore, a list of techniques is created, that could lead to an improvement of the quality of website testing results. Recapitulatory, concrete proposals for continuative research are presented.

Keywords: Accessibility, WCAG, Accessible websites

1 Introduction

Inclusion, the process of allowing people with disabilities to participate in normal life, has been an ongoing process for many years. One side-effect of this process is, that a growing number of handicapped people starts using the internet as information source or for activities like online-shopping or partaking in social media.
Unfortunately, the web lost its capabilities concerning cross-device usage, in the course of development towards 'Web 2.0' (def. [O'R06]). The approaches to supply the user with websites providing Software as a Service (SaaS) or Desktop as a Service (DaaS) is often based on heavy usage of client-side scripts, which prevents assistive software from being able to gather all information on a web page.
In 1997, the W3C founded a group to focus on accessible web pages, the WAI. This group released the first version of their Guidelines in 1999, the second one in 2008. The currently valid release WCAG 2.0 was adopted by the International Organization for Standardization (ISO) in 2012 [ISO12b]. The WCAG 2.0 will be referred to as the WCAG below. Notable changes are the switchover from backward to forward compatibility and the new structure of the guidelines themselves.
Unfortunately, the WCAG do not include guidelines to make websites accessible for people with cognitive impediments, like text complexity or readability of text. Neither do they

[1] Technische Hochschule Ingolstadt, Fakultät Elektrotechnik und Informatik, Esplanade 10, 85049 Ingolstadt, mah0899@thi.de
[2] The WCAG are published online: [CRCV08]

include any testing methods for their guidelines.
To enable assistive technologies to recognize the semantics of Hypertext Markup Language (HTML) elements more precisely, the WAI introduced the Accessible Rich Internet Applications (ARIA) [CC14], which primarily provides a huge list of roles an element can take. The ARIA Standard is a completed W3C Recommendation since March 2014 and thus not fully implemented in current browsers. In fact, the major browsers provide all partial support [Can15].

2 Implementation in Germany

Mindless of the fragmentariness of the WCAG, the German legislative enacted the 'accessible information technology act', *Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (BITV)*, to regulate the accessible presentation of websites of public institutions. Initiated by the *DIAS GmbH* (`http://www.dias.de/`), the Project Barrierefrei informieren und kommunizieren (BIK) provides a testing framework to validate the compliance of websites with the BITV. The testing procedures are mostly rather complicated to perform and focus on the BITV very strictly.
After all the derival of guidelines, the BIK-test has lost the focus to test accessibility for people with disabilities in favor to comply with the applicable law.

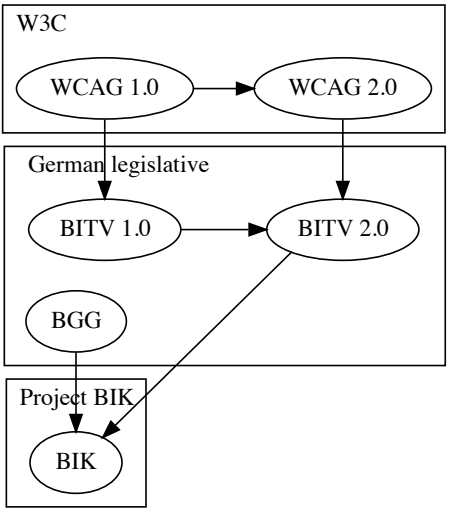


Fig. 1: Development of the BIK-test.

3 Contradictions and obsolete guidelines

In addition to the unappreciated groups of handicapped people, some of the guidelines rely on obsolete perceptions. Other rules do not scrutinize the reasons in a penetrative manner,

so that the levels of conformance do not have a realistic basis. The most demonstrative examples are examined below in a critical way.

3.1 Flashes

In *WCAG 2.3: Seizures*, the *Three Flashes or Below Threshold* is introduced. This limitation is based on data by *Harding, G.F.A* from 1975, but used in the work of *Harding*, that was used to create the WCAG (cf. [BEG+02]). Numerous papers written by other authors before 1975 reference data collected by *Gibbs* in 1935 (cf. [GDL35]), e.g. [Qua39, Gib49, LHF59].
It is not sure, that these values are still valid, since the influence of media to the people and society has boomed since then. As measured by the media-usage of 6-12 year olds[3], people use more digital media today and start using it at younger age. The cerebral habituation has never been investigated, so it is not clear, if a brain, that is photosensitive may acclimatize to flickering over the years. In 1975, when the data was collected, computer monitors were all based on the Cathode Ray Tube (CRT) technology. *"Flat panel displays"* were introduced in the 90's (cf. [Bou95]) and one of their advantages is a non-flickering image, which changes the way the photoreceptors are stimulated. This means, the same type of flickering may cause different impressions on Thin-film transistor (TFT) than on CRT displays, as the frequency of the flickering may also cause interferences with the flickering of the CRT and influence the experienced frequency. TFT-monitors use other techniques and therefore refresh rates between 60 and 75 Hz as confirmed by *Grundig*.
But the flashing frequency is not the only epileptic trigger. Epilepsy patients often show signs of an oncoming seizure hours before [LE02]. In this state, the patient is much more sensible to any kind of stimuli, so they react to flickering on lower frequency. Additionally, the seizure risk is dependent on other values, like the flashing percentage of the visual field and the relative luminance of the flash [CRCV08]. In summary, one needs to question if the current threshold is still a realistic value. As flashes can be counted as animations[4], the usage of flashes contradict the *WCAG 2.2.2: Pause, Stop, Hide*, as the user must be able to turn animations off. This topic will be discussed in detail later.
Another kind of flickering not discussed in the WCAG is the usage of multicolor background images with high contrast. When the page needs to be scrolled, the contrast changes with the moved image. When using a checkerboard pattern with fixed field sizes, the scrolling in different speeds leads to a different frequency of a full change in contrast. The frequency is calculated with the scrolling speed v ($\frac{\text{px}}{\text{s}}$)[5] and the grid size l (px) as in (1):

$$f = \frac{v}{2 \cdot l}(\text{Hz}) \tag{1}$$

The following shows some sample calculations derived from data using a screen with a resolution of 1920 by 1080 pixels. To use scrolling speed, that is near realistic, *Facebook* (`http://www.facebook.com`) was chosen, as a highly frequented page, where scrolling

[3] *KIM-Studie* 2014: [FPR14]
[4] A flash is a "rapidly changing image sequence" [CRCV08, #general-thresholddef].
[5] px: Definition found at [JEL13, #absolute-lengths]

is essential. A scrolling speed of approximately 800 $\frac{\text{px}}{\text{s}}$ was found. As shown in [HJ94, p. 87], 96% of photosensitive people react to a frequency span between 15 and 20 Hz. Depending on the personal scrolling speed, high-contrast patterns can produce flickering in very critical frequencies, as shown examplarily in figure 2.

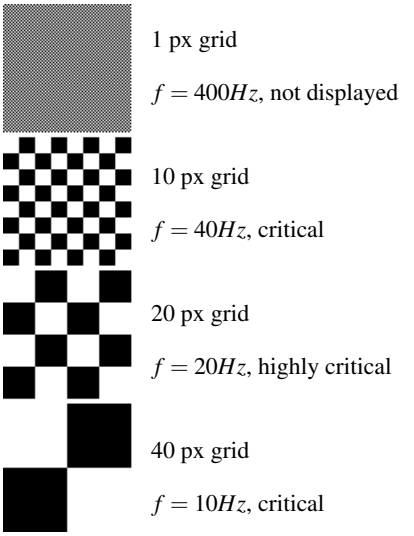


Fig. 2: Usage of a checkerboard of different grid sizes at a scrolling speed of 800 $\frac{\text{px}}{\text{s}}$.

3.2 Cognitive Load Theory

Before any information can partake in the cognition process, the human brain must perform complex perceptive and filtering tasks. This process is illustrated by the Model Human Processor (MHP), "an engineering model of human performance" [Gei06, p. 30ff]. The stimuli encounter the sensory organs and are converted into nerval impulses. These impulses are processed by the perceptual processor and buffered into their respective visual or auditory image store in the working memory. Only now is the cognition process performed by the cognitive processor. The amount of data that needs to be evaluated to participate in this process is a significant factor of the duration until information is processed.
The *Load theory* by Lavie [LHdFV04] claims that every person owns a limited perceptual buffer. This means that any cognition process uses a more or less vast amount of the available capacity, depending on the complexity of the task. When the provided information contains additional stimuli carrying other information, the perception will need more time. The cognition of information is also harder for the person [Gol15, p. 142f].
People with cognitive disorders feature shortcomings in one or more parts of the cognitive system [Pfl91, p. 166]. This leads to either a smaller cognitive capacity or a higher load. On both sides, there is a need to provide simple and clear structured information in order to allow these people to perceive it.
The Guidelines *WCAG 1.4.2 Audio Control* and *WCAG 2.2.2 Pause, Stop, Hide* deal with

time based media, like audio, video and animations. The guidelines suggest, that for automatically started content of these types, a mechanism to disable it needs to be provided. The problem is, that when distractive content is started automatically, searching the mechanism to turn it off is a cognitive challenge, that people that depend on turning it off may not be able to find. This means, that these rules do not meet their aims since they should rather require that content must not be distractive at all.

3.3 Wrong interpretation

Many websites that claim to be accessible provide a method to enlarge the content of the website using buttons. This seems to be a misinterpretation of *WCAG 1.4.4 Resize text* and *WCAG 1.4.8 Visual presentation; Nr. 5*, which instructs that websites "can be resized without assistive technology up to 200 percent without loss of content or functionality." Every major browser[6] supports this technique by default, so there is no need to provide additional methods to zoom in the page. Rather important is to have the layout of the page enabled to be zoomed without changing the usability or the contents.
The derivate towards the BITV lead to other misinterpretations. After BITV 2.0 §3(2), only an overview of the website needs to be fully accessible using the German Sign Language (gsg) [ISO10] and *Leichte Sprache* (`http://www.leichtesprache.org/`), an approach to create rules to provide a maximum understandable language. This means, the legal requirements can be complied without having to ensure the accessibility of the vast part of the provided information.

4 Shortcomings in the WCAG

4.1 Cognitive and cerebral handicaps

Since the WCAG focuses primarily on the semantics and visual scopes of websites, the understandability of content is not respected in the same manner. This might be due to the complex calculations of text readability, if it is possible to define it at all. *WCAG 3.1.5 Reading Level* proposes, that all texts should be readable for any person with a "lower secondary education level" as described by the *UNESCO* [UNE97]. This might often be difficult to define by website editors and does cost high effort. So, a possible approach towards automatization of determination of text reading ease must be found.
Although multiple measuring techniques for text complexity exist, most of these were created with the target to define the minimal level of education the reader needs to have achieved to understand the text, like the *Flesch-Reading-Ease* [Bau03, p. 35f], [Fle48] or the *Gunning-Fog-Index* [Gun68]. In his master thesis (cf. [Ott09]), Niels Ott suggests techniques to describe text complexity using Lexical Frequency Profiles (LFPs), which were introduced by Laufer and Nation [LN95]. Ott further describes a process, to allow

[6] Regarding *Google Chrome*, *Microsoft Internet Explorer*, *Mozilla Firefox*, *Apple Safari* and *Opera* according to [Sta15]

automated computation of LFPs. This could be an approach to retrieve readability data from web page contents, and to define thresholds for cognitive disabled people.
To implement this kind of guideline, it must be determined, that not only presented text must follow the rules, but also auxiliary ones, like alternative descriptions for images and other objects or definitions for expressions.

4.2 Readability, legibility and comprehensibility

Most of the information on the internet is in text form. In *WCAG 1.4.8 Visual presentation*, some guidelines are presented to improve legibility and comprehensibility. Since readability refers to the perceptibility of single characters and legibility to the perceptibility of an amount of coherent characters [DIN98], it is necessary to have text readable in order for it to be legible. In DIN 1450 [DIN13], characteristics of fonts and text presentation are defined. The WCAG provide rules for text design concerning contrast (*WCAG 1.4.6 Contrast (Enhanced)*), line spacing, alignment and scaling, as well as text authoring rules like maximum characters per line. Unfortunately, there is no rule for minimal size of continuous text, since there are several standards defined for minimum font size.
The German project *Leichte Sprache* has created a set of rules to make text understandable to a maximum number of people by defining rules for readability, legibility and comprehensibility [Spr13]. Their suggested minimum font size is 14 pt, which equals $1^1/_8$ em.
The standard minimum font size according to ISO 9241-302:2008 [ISO12a] is 16 arc minutes, giving a font size of 1.1 mm or 4 px.
In order to create websites, that change their layout appropriate for any end device, the Responsive Web Design (RWD) approach recommends to use a minimum font size of 1 em [Ert13, p. 133ff].
Since text can be resized by the browser, there is no need to provide a very huge font. But in order to implement a web site that is state of the art, a font size of 1 em should at least be used.
Leichte Spache provides a very good set of rules, that could be adopted into a WCAG successor, like avoiding fonts with serifs or italic displayed text. Also, the authoring rules provide outstanding approaches to achieve easily understandable text.

4.3 Partitioning classes of accessibility

The WCAG approach to classify the accessibility of websites groups the compliance with the guidelines in three conformance levels:

A	Minimum
AA	Mediocre
AAA	Maximum

To achieve any of the conformance levels it is necessary, that, according to [CRCV08], “the Web page satisfies all the [...] Success Criteria, or a conforming alternate version is

provided." This means, there is no distinction between guidelines, that provide users with a better usage of assistive devices and the direct accessibility, like easy reading or definitions of technical expressions. A better approach would be, to partition the two classes and evaluate them separately. A short investigation whether the guidelines affect the information retrieval for people or machines yields, that 25 of them focus on human interaction, 13 on semantic readability for assistive technologies and 22 on both. This means, a website could achieve an AA-Level without being usable for people with cognitive disabilities just because of semantically correct content.
To allow a distinction if a website is good for assistive technologies or for human cognition, there is a need to implement two different levels of accessibility achieved. This would allow a more explicit definition of a page's accessibility.

5 Proposed changes to the WCAG

5.1 Flashes

As already pointed out, the flash frequency threshold needs to be reviewed. As an addition to the *WCAG 2.3: Seizures*, the need for a new success criteria is revealed, that demands the investigation of images, that fill "25% of any 10 degree visual field on the screen" for the way they change contrast when scrolled at a distinct scrolling speed. The biggest problem is that the scrolling speed is not only different for every person, but also for every website. While websites with a lot of text will be scrolled slower in order to follow the text, pages with image content are scrolled way faster to "jump" from one image to the next one. To achieve to measure the effective change of contrast, the contrast changing frequency on TFTs needs to be investigated.
In addition, a study must verify that any photosensitive person who takes part is not in "aura" status (cf. [SLN+95]) in the moment of testing. This would lead to falsified results.

5.2 Automatic determination of text complexity

Testing texts on websites manually would enormously raise the testing effort. So it is mandatory to use an automated testing of texts. A good entry point is the combination of different text-complexity-measuring systems described by *Niels Ott* [Ott09]. When a system based on this thesis can be realized, it can provide a very distinct value to describe the reading ease. If it should turn out that it is not possible to create an automated test, the WCAG still need to specify a more concrete value measured with one specified method for the determination of understandability.

6 Tools proposed for web page testing

To allow anyone to test websites for accessibility, the whole process must depend on freeware or open source tools that can be run on any major operation system. The suggested

approach is using an open source browser and a maximum number of browser plug-ins, that support the testing procedure.
The selection of the browser is done by comparing the HTML5-compliance of current browsers, as *WCAG 4.1 Compatible* instructs to "Maximize compatibility with current and future user agents". The website `https://html5test.com/` provides a testing framework to show the HTML5-compliance of browsers. This page provides test results for the most current browsers with their distinct results. The results provided by *html5test* have been compared with the ones we determined ourselves in table 1. Since neither *Microsoft*

	Provided result	Determined
Google Chrome	39; 501	42; 523[7]
Opera	26; 497	29; 519
Mozilla Firefox	35; 449	37; 449
Apple Safari	8.0; 396	8.0.4; 396
Microsoft Internet Explorer	11; 336 Preview[8]; 343	N/A

Tab. 1: Comparison of HTML5-compliance of mainstream browsers by version and score (Max. 555); tested on *Mac OS X 10.10.2*.

Internet Explorer nor *Apple Safari* provide native Linux versions, the browsers can not be used for testing on Linux systems and are therefore discouraged to use for website testing. This means, there might be minor differences between the presentation in these browsers, but when development tracks the compliance with standards, testing with the best complying browser means testing for forward compatibility.
The results for Opera and Chrome respectively Chromium, its open source fork, are rather close, as both browsers could be used for website testing. Both browsers are based on *WebKit*, an open source HTML rendering engine. Since *Google Chrome* is the most used browser focus is set on it.
The first add-on, which allows several tests is *WebDeveloper* (`http://chrispederick.com/work/web-developer/`). This add-on provides a huge toolkit for website examination and allows to verify a vast part of the WCAG, mostly the semantic parts.
To check the contrasts of anything displayed on the website, a very handy tool comes with the *Color Contrast Analyzer* (`http://accessibility.oit.ncsu.edu/tools/color-contrast-chrome/`). This add-on scans the page for any visible content, that provides at least a contrast level of 3:1, 4.5:1 and 7:1 for all contrast thresholds mentioned in the WCAG.
Another test for visual impaired people is testing for types of Color Vision Deficiency (CVD). A very good add-on for *Google Chrome* is provided through *Spectrum* (`http://lvivski.com/spectrum/`), a simple tool, that changes the colors of any website according to what they look when perceived with a distinct CVD.
To perform markup analysis, a great assistance is provided by *Google Chrome* itself by providing the *Chrome Developer Tools* [Goo15], that allow live manipulation of the Document Object Model (DOM) and Cascading Stylesheets (CSS), as well as tools to measure loading times and other features. Finding HTML elements, that do not comply with standards is easily possible, as well as finding out, how the browser processes corrupt HTML. Checking any page for tags and attributes is much simpler than searching in plain source

code.
Since most guidelines must be evaluated by a human, there may exist more add-ons to support the human-driven testing procedure but the operation purpose depends on the personal knowledge of the testing person.

6.1 Required Hardware

To perform a realistic test on "flashes", the contrast change by time, additional hardware is needed. Since there is no proposed testing procedure in the WCAG or the BIK-test, a possibility to perform correct tests had to be discovered.
The company *PCO* (`http://www.pco.de/`), developer of "highend scientific camera systems", suggests testing with a specifically programmed camera. They suggest testing with a low resolution camera with high color depth and a frame rate with at least 120 frames per second, doubling the TFTs refresh rate. Using their Application Programming Interface (API), an automatic determination of contrast changes in any visual field on the screen could be performed and alerted.
Though it is contradictory to develop a device, that might not be affordable for every person that wants to check a website, it is still possible to estimate the compliance with the thresholds, like the BIK-test suggests [DIA11], but allows a more exact examination in case of further investigation.

7 Conclusion and Outlook

Since the WCAG are still very fragmentary, and a lot of guidelines can not be tested using currently available tools, a lot of research needs still to be done to allow an objective view on accessibility of web pages. There is also a need to change the regrettable loss of focus on people that feature other types of handicaps than visual impairments. The approach towards the creation of a global set of guidelines is done rather good by the WCAG, but the first version had more practical definitions, especially the focus on HTML features. The WCAG 2.0 guidelines are held very rough and leave therefore space for interpretation that may lead to both good and bad results. The German implementation exposes even more gaps. Starting from estimation of compliance instead of providing specific tools, to testing of obsolete techniques that should not be used anyway.
The next steps towards better testing of websites that can be named from the conclusions of this paper are surely not the only fields where research is needed. Nevertheless, some suggestions to achieve the required techniques to fulfill the described shortcomings are made below.

7.1 Automatic semantics testing

Some testing steps are easier to achieve with automatic processing than manually by an untrained person. Since most semantic rules concern nesting and attributes of HTML ele-

ments, a vast part of the WCAG can be tested by analyzing the HTML document. Fortunately, there are several techniques to dissect HTML documents, like the DOM [HRC+04]. This allows to perform analysis on markup concerning e.g. correct markup usage or avoidance of obsolete HTML elements. In addition, the basic tests for syntax validity can be verified using the W3C validator (`http://validator.w3.org/`). Most tests must however be done manually at the moment, since they require mental work and therefore need to be rated by a human.

7.2 Automatic front-end testing

Since verifications of questions like "does the alternative text for this image make sense?" need to be answered by a person, automatization of the testing is hard to achieve. For most parts add-ons can help answering, but the determination of results must be done by a human. But automatization can be achieved step by step. An approach to enable automatic testing for not only the static parts of a website, but also dynamic content, could be achieved by using front-end testing tools. One developer of *Mozilla* created a list of front-end testing tools for websites [Dub12]. The steps preformed by the testing tool are defined in configuration scripts, that advise the framework what to do. Some tools even simulate mouse interaction.
When combined with a system that acts like a web crawler for the one page to test, this system could dynamically create the configuration for the testing framework and create a flow to check every page. If this can be achieved, the particular content tests can be executed automatically for every page or after each content-changing event. As the development of techniques for miscellaneous types of content progresses, testing may be integrated into this workflow.
With this approach, a maximum automatization of the entire testing procedure can be achieved over time.

7.3 Development of a text measuring engine

The most complex step, but also that one, that could maximize the amount of test automatization is the development of a system, that can perceive and evaluate text and afterwards rate its comprehensibility. This means immense effort to achieve, since the system must contain accumulated knowledge about the grammatics and a huge dictionary of every single language aimed to work with. This could force the usage of some kind of distributed database that is aware of correlations between the words. This knowledge of relations is a characteristic feature of graph databases [RWE13, p. 4], and thus an approach to create a linguistic processor on top of a graph database system seems to be very very promising.
In their paper [MP14], *Klaus Miesenberger* and *Andrea Petz* suggest some approaches to achieve a kind of toolkit, that can be used to verify easy readability using "linguistics and language technologies". They suggest, that the whole process of content creation should be accompanied by tools to analyze, check and translate text to simpler language. This approach could avoid, that very complex texts don't even go online if tools are integrated into the authoring workflow.

Acknowledgment

This work was supported by *Ferdinand Haimerl* with continuous professional input and frequent feedback.

List of acronyms

API Application Programming Interface

ARIA Accessible Rich Internet Applications

BIK Barrierefrei informieren und kommunizieren

BITV Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz

CRT Cathode Ray Tube

CSS Cascading Stylesheets

CVD Color Vision Deficiency

DaaS Desktop as a Service

DOM Document Object Model

gsg German Sign Language

HTML Hypertext Markup Language

ISO International Organization for Standardization

LFP Lexical Frequency Profile

MHP Model Human Processor

RWD Responsive Web Design

SaaS Software as a Service

TFT Thin-film transistor

W3C World Wide Web Consortium

WAI Web Accessibility Initiative

WCAG Web Content Accessibility Guidelines

References

[Bau03] Andreas Baumert. *Professionelles Texten. Tipps und Techniken für den Berufsalltag.* Beck Juristischer Verlag, 2003.

[BEG+02] CD Binnie, J Emmett, P Gardiner, GFA Harding, D Harrison, and AJ Wilkins. Characterizing the flashing television images that precipitate seizures. *SMPTE Motion Imaging Journal*, 111(6-7):323–329, 2002.

[Bou95] William R. Boulton. Note on the Liquid Crystal Display Industry, 1995.

[Can15] Can I use. WAI-ARIA Accessibility features, April 2015.

[CC14] Michael Cooper and James Craig. Accessible Rich Internet Applications (WAI-ARIA) 1.0. W3C recommendation, W3C, March 2014. http://www.w3.org/TR/2014/REC-wai-aria-20140320/.

[CRCV08] Ben Caldwell, Loretta Guarino Reid, Michael Cooper, and Gregg Vanderheiden. Web Content Accessibility Guidelines (WCAG) 2.0. W3C recommendation, W3C, December 2008. http://www.w3.org/TR/2008/REC-WCAG20-20081211/.

[DIA11] DIAS GmbH. Prfschritt 2.3.1a: Verzicht auf flackern, October 2011.

[DIN98] DIN. Begriffe der physiologischen Optik. Norm DIN 5340, Deutsches Institut für Normung, Berlin, Germany, April 1998.

[DIN13] DIN. Schriften - Leserlichkeit. Norm DIN 1450, Deutsches Institut für Normung, Berlin, Germany, April 2013.

[Dub12] K. Dubost. Compatibility/web testing: Automated testing for web applications and web sites, September 2012.

[Ert13] Kai Laborenz Andrea Ertel. *Responsive Webdesign*. Galileo Press GmbH, 2013.

[Fle48] Rudolph Flesch. A new readability yardstick. *Journal of applied psychology*, 32(3):221, 1948.

[FPR14] S Feierabend, T. Plankenhorn, and T. Rathgeb. KIM-Studie 2014: Kinder und Medien, Computer und Internet. Basisuntersuchung zum Medienumgang 6- bis 13-Jähriger in Deutschland, February 2014.

[GDL35] FA Gibbs, H Davis, and WG Lennox. The EEG in epilepsy and in the impaired states of consciousness. *Arch Neurol Psychiatry*, 34:1133, 1935.

[Gei06] Jürgen Geisler. *Leistung des Menschen am Bildschirmarbeitsplatz: das Kurzzeitgedächtnis als Schranke menschlicher Belastbarkeit in der Konkurrenz von Arbeitsaufgabe und Systembedienung*. Universitätsverlag Karlsruhe, 2006.

[Gib49] Frederic A Gibbs. Der gegenwärtige Stand der klinischen Elektrencephalographie. *European Archives of Psychiatry and Clinical Neuroscience*, 183(1):2–11, 1949.

[Gol15] E. Bruce Goldstein. *Wahrnehmungspsychologie: Der Grundkurs (German Edition)*. Springer, 2015.

[Goo15] Google. Chrome devtools overview, May 2015.

[Gun68] Robert Gunning. *The Technique of Clear Writing.* Mcgraw-Hill, 1968.

[HJ94] Graham FA Harding and Peter M Jeavons. *Photosensitive epilepsy*. Number 133. Cambridge University Press, 1994.

[HRC+04] Philippe Le Hégaret, Jonathan Robie, Mike Champion, Lauren Wood, Steven B Byrne, Gavin Nicol, and Arnaud Le Hors. Document Object Model (DOM) Level 3 Core Specification. W3C recommendation, W3C, April 2004. http://www.w3.org/TR/2004/REC-DOM-Level-3-Core-20040407.

[ISO10] ISO. Codes for the representation of names of languages – Part 3: Alpha-3 code for comprehensive coverage of languages. ISO ISO/IEC 4639-3:2007, International Organization for Standardization, Geneva, Switzerland, July 2010.

[ISO12a] ISO. Ergonomics of human-system interaction – Part 302: Terminology for electronic visual displays. ISO ISO 9241-302:2008, International Organization for Standardization, Geneva, Switzerland, March 2012.

[ISO12b] ISO. Information technology – W3C Web Content Accessibility Guidelines (WCAG) 2.0. ISO ISO/IEC 40500:2012, International Organization for Standardization, Geneva, Switzerland, October 2012.

[JEL13] Tab Atkins Jr., Elika Etemad, and Håkon Wium Lie. CSS Values and Units Module Level 3. Candidate recommendation, W3C, July 2013. http://www.w3.org/TR/2013/CR-css3-values-20130730/.

[LE02] Brian Litt and Javier Echauz. Prediction of epileptic seizures. *The Lancet Neurology*, 1(1):22–30, 2002.

[LHdFV04] Nilli Lavie, Aleksandra Hirst, Jan W de Fockert, and Essi Viding. Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, 133(3):339, 2004.

[LHF59] A Lundervold, GF Henriksen, and L Fegersten. The spike and wave complex; a clinical correlation. *Electroencephalography and clinical neurophysiology*, 11(1):13–22, 1959.

[LN95] Batia Laufer and Paul Nation. Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*, 16(3):307–322, 1995.

[MP14] Klaus Miesenberger and Andrea Petz. "Easy-to-Read on the Web": State of the Art and Needed Research. In Klaus Miesenberger, Deborah Fels, and Dominique Archambauldt, editors, *Computers Helping People With Special Needs*, pages 161–168. Springer, July 2014.

[O'R06] Tim O'Reilly. Web 2.0 Compact Definition: Trying Again, December 2006.

[Ott09] Niels Ott. Information retrieval for language learning: An exploration of text difficulty measures. Master's thesis, Universität Tübingen, Tübingen, Germany, 2009.

[Pfl91] Leander Pflüger. *Neurogene Entwicklungsstörungen: eine Einführung für Sonder- und Heilpädagogen*. Reinhardt, München, 1991.

[Qua39] J. H. Quastel. Respiration in the central nervous system. *Physiological Reviews*, 19:135–138, April 1939.

[RWE13] Ian Robinson, Jim Webber, and Emil Eifrem. *Graph Databases*. O'Reilly Media, 2013.

[SLN+95] R Schulz, HO Lüders, S Noachtar, T May, A Sakamoto, H Holthausen, and P Wolf. Amnesia of the epileptic aura. *Neurology*, 45(2):231–235, 1995.

[Spr13] Netzwerk Leichte Sprache. Die Regeln für Leichte Sprache. 2013. http://www.leichtesprache.org/downloads/Regeln%20fuer%20Leichte%20Sprache.pdf.

[Sta15] StatCounter. Statcounter global stats: Top 5 desktop, tablet & console browsers from apr 2014 to apr 2015, April 2015.

[UNE97] UNESCO. International Standard Classification of Education, 1997.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Fehlertolerante und energieeffiziente eingebettete Systeme: Methoden und Anwendungen

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Fehlertolerante und energieeffiziente eingebettete Systeme: Methoden und Anwendungen

Mario Schölzel[1], Heinrich Theodor Vierhaus[2]

Abstract: Die zunehmende Miniaturisierung bei der Fertigung digitaler Schaltungen ermöglicht die Herstellung immer kleinerer, kostengünstigerer und trotzdem leistungsfähigerer Systeme. Diese Trends haben zu einer enormen Verbreitung solcher Systeme in fast allen Bereichen des täglichen Lebens geführt, beginnend bei der Automobilelektronik über die Avionic bis hin zur Steuerung, Regelung und Überwachung von Maschinen und Geräten in Industrieanlagen, Häusern und anderen Infrastrukturbereichen. Durch diese Durchdringung unserer Umgebung mit eingebetteten Systemen, wird die Abhängigkeit von ihrer korrekten Funktion in Zukunft noch stark zunehmen. Auf der anderen Seite führt die Miniaturisierung bei der Fertigung aber auch zu einer steigenden Anfälligkeit der Systeme gegenüber transienten, intermittierenden und permanenten Hardwarefehlern, die im laufenden Betrieb auftreten. Fehlertoleranztechniken bieten eine Möglichkeit auch unter solchen Bedingungen die korrekte Funktion für einen langen Zeitraum zu ermöglichen. Solche Maßnahmen erfordern jedoch Redundanz, die den Energieverbrauch deutlich erhöhen kann. Das ist insbesondere für eingebettete Systeme problematisch deren Energiebudget stark limitiert ist. Es ist deshalb von zunehmender Bedeutung, bei der Entwicklung von Fehlertoleranztechniken für eingebettete Systeme besonders auf den sparsamen Umgang mit vorhandenen Energieressourcen zu achten, um einen zuverlässigen Langzeitbetrieb zu erlauben.
Durch drei eingeladene Tutorials wird zunächst eine Einführung in die Thematik und etablierte Techniken gegeben. Aktuelle Methoden für die Realisierung zuverlässiger und energieeffizienter eingebetteter Systeme werden anschließend durch Präsentation der begutachteten Workshop-beiträge vorgestellt und diskutiert.

Programmkomitee

Dr. J. Alt	(Intel Mobile Communications GmbH)
Dr. M. Dietrich	(Fraunhofer IIS, Institutsteil Entwurfsautomatisierung EAS)
Prof. Dr. M Gössel	(Universität Potsdam)
Prof. Dr. S. Hellebrand	(Universität Paderborn)
Dr. T. Koal	(BTU Cottbus-Senftenberg)
Prof. Dr. R. Kraemer	(IHP Frankfurt (Oder) und BTU Cottbus-Senftenberg)
Prof. Dr. P. Langendörfer	(IHP Frankfurt (Oder) und BTU Cottbus-Senftenberg)
Dr. T. Preußer	(TU Dresden)
Prof. Dr. B. Schnor	(Universität Potsdam)
Prof. Dr. M. Schölzel	(IHP Frankfurt (Oder) und Universität Potsdam)
Prof. Dr. H.T. Vierhaus	(BTU Cottbus-Senftenberg)
Prof. Dr. H. J. Wunderlich	(Universität Stuttgart)

[1] IHP GmbH Frankfurt (Oder) und Universität Potsdam, schoelzel@ihp-microelectronics.com
[2] BTU Cottbus-Senftenberg, Institut für Informatik, Heinrich.Vierhaus@b-tu.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Low Power Design Techniques

Goran Panić[1]

Abstract:

The demand for portable electronic devices that offer greater functionality and performance at lower costs and smaller sizes has increased rapidly. This market trend is driving the need for efficient System-on-Chip (SoC) designs, where the power arises as the one of the biggest problems. The microprocessor design has traditionally focused on dynamic power (active power) consumption as the limiting factor in system integration. As technology has shrunk to 90 nm and below, static power (leakage power) is posing new challenges to low power design.

Historically, CMOS technology has dissipated much less power than earlier technologies such as transistor-transistor and emitter-coupled logic. In fact, when not switching, CMOS transistors consumed negligible power. However, with the increase in device speed and chip density, the power of CMOS increased dramatically. According to technology trends, the dynamic power per device decreases over the time. However, if it is assumed that the number of on-chip devices doubles every two years, total dynamic power increases on per-chip basis. Additionally, the shrinking of feature sizes makes static power dissipation grows exponentially. Consequently, the packaging and cooling costs as well as the limited power capacity of batteries become unsustainable.

The reduction of power growth below the predicted numbers has become one of the most important tasks for designers. To cope with the power challenges, a number of advanced power saving techniques have been developed that can efficiently target both static and dynamic power loss. The existing and future trends in low power design show that efficient power management is greatly determined by design decisions made in early phase of system design. Thus, the selection of power saving strategy becomes one of the most important challenges for designers.

This presentation starts by covering basic sources of power consumption in CMOS, including both static and dynamic power loss. Furthermore, the presentation gives a detailed overview on existing low power techniques including standard techniques developed for mature technologies, process-level techniques, and advanced low power techniques used in deep-submicron CMOS. Also, some attention is given to the selection of power saving strategy for design and early power estimation. Finally, future trends in low power design are discussed as well.

Keywords: low power, system-on-chip, dynamic power, static power,

[1] IHP, Im Technologiepark 25, 15236 Frankfurt (Oder), Germany, panic@ihp-microelectronics.com

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Reducing Power Consumption in Fault Tolerant ASICs

Milos Krstic[1]

Abstract: Fault tolerance techniques have been traditionally used for specific applications such as space, avionic, or nuclear where the reliability is of utmost importance. Moreover, nowadays the reliability of nanoscale CMOS processes is reduced, and the power optimization goes beyond the standard worst-case boundaries. As a consequence, fault tolerance became the important factor also for the main stream industry. Handling faults is connected with some sort of redundancy either in hardware, timing, information, or by combination of previously defined methods. The immediate consequence of the redundancy is increased power consumption. As an example, the traditional N-modular redundancy techniques, such as TMR (triple modular redundancy) end up in the power overhead of about 300-400% of original non-protected system. In order to limit the overhead the additional methods and techniques need to be applied. In this presentation the relevant state-of-the-art methods for reducing power consumption in fault tolerant chips will be presented. This will include the global overview of the traditional methods (such as improvements of TMR/DMR schemes) and novel methods for low-power ASICs (such as RAZOR). Moreover, the methods at different abstraction layers proposed at IHP for handling the power optimization in fault-tolerant ASICs will be presented. At the RTL level, selective fault tolerance can be applied limiting overhead by 50% with still protecting 20% of the inputs. Additionally, for the memories the overhead can be reduced by utilizing specific error correction codes. Finally, the fault tolerance can be introduced on application demand, dramatically reducing the power consumption when it is not needed. This can be performed on the pipeline level or even at the level of multi-processors. Such an example will be shown and IHP's FMP adaptive multi-core platform will be presented.

Keywords: fault tolerance, TMR, low power

[1] IHP, Im Technologiepark 25, 15236 Frankfurt (Oder), Germany, krstic@ihp-microelectronics.com

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Fehler vermeiden oder tolerieren? Methoden zur Berücksichtigung der Zuverlässigkeit im Entwurf integrierter Schaltungen

Roland Jancke[1], Christoph Sohrmann[2], Kay-Uwe Giering[3]

Abstract: Integrierte Halbleiterelektronik wird zukünftig immer mehr Lebensbereiche durchdringen. Für hohe Leistung bei geringstem Energieverbrauch und minimalen Abmessungen kommen neueste Technologie-Generationen zum Einsatz. Die Anwendungen reichen vom Consumer-Bereich bis hin zu sicherheitskritischen Bereichen wie Automotive und Industrie 4.0.

Allerdings bringen neue Halbleiter-Technologien auch neue Materialien und neue physikalische Effekte mit sich, die die Zuverlässigkeit der Elektronik und des Gesamtsystems beeinträchtigen. Es wird daher immer wichtiger, die Auswirkungen von zuverlässigkeitsrelevanten Effekten einschätzen und im Entwurf integrierter Elektronik berücksichtigen zu können.

Im Beitrag werden grundsätzliche Herangehensweisen zur Absicherung der Elektronik-Zuverlässigkeit unterschieden. Im Mittelpunkt der Betrachtungen stehen intrinsische Degradationseffekte, die zum Ausfall von Elektronik über der Zeit führen können. Es werden Verfahren vorgestellt, die es gestatten, solche Effekte bereits im IC-Entwurf zu berücksichtigen und dadurch alterungsbedingte Ausfälle zu vermeiden.

Die Methoden beruhen auf der Kenntnis der zugrundeliegenden physikalischen Degradationseffekte, wie sie in der Literatur beschrieben und von den Halbleiterherstellern charakterisiert werden. Heute werden aus diesen Daten bereits wichtige Erkenntnisse zur Entwicklung und Qualifizierung der Technologie gezogen. Der vorgestellte neue Ansatz nutzt die Charakterisierungsergebnisse zur Erstellung von simulationsfähigen Modellen und als Randbedingung für den Entwurf zuverlässiger Schaltungen. Damit lassen sich quantitative Aussagen zum Verhalten einer Schaltung für eine prognostizierte Lebensdauer ableiten sowie die Zuverlässigkeit als eine weitere Randbedingung bei der Optimierung von IC-Entwürfen verwenden. Ziel ist die Absicherung von Zuverlässigkeitsvorgaben bei der Entwicklung elektronischer Schaltungen und Komponenten für sicherheitskritische Anwendungen.

[1] Fraunhofer Institut für Integrierte Schaltungen IIS, Institutsteil Entwurfsautomatisierung EAS Dresden
[2] Fraunhofer Institut für Integrierte Schaltungen IIS, Institutsteil Entwurfsautomatisierung EAS Dresden
[3] Fraunhofer Institut für Integrierte Schaltungen IIS, Institutsteil Entwurfsautomatisierung EAS Dresden

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Kompression von Tracedaten auf Bitebene basierend auf einem LZ77-Wörterbuchansatz

Kai-Uwe Irrgang[1] Thomas B. Preußer[2] Rainer G. Spallek[2]

Kurzfassung: Online-Tracing ist ein leistungsfähiger Ansatz für das Monitoring und Debugging von eingebetteten Prozessoren. Die aufgezeichneten Tracedaten ermöglichen nachfolgend selbst aufwendige Analysen von Systemläufen. Während des Tracings wird ein sehr hohes Datenvolumen erzeugt. Dieses stellt eine große Herausforderung für Speichersysteme und Kommunikationsschnittstellen dar. Eine unmittelbare, möglichst quellnahe und signifikante Datenreduktion in Echtzeit ist außerordentlich wichtig. Die beiden Hauptstrategien dafür sind die Relevanzfilterung und die Kompression der Tracedaten.

Beim Instruktionstrace existieren mehrere effiziente Lösungen für die Kompression von Zieladressen ausgeführter Verzweigungsbefehle. Moderne Befehlssatzarchitekturen erweitern jedoch die Möglichkeit der bedingten Befehlsausführung über Verzweigungsbefehle hinaus auf mehr oder gar alle Befehlsgruppen. Dies führt zu umfangreichen Strömen von Ausführungsbits, die anzeigen, ob die jeweiligen bedingten Instruktionen ausgeführt wurden oder nicht. Dieser Beitrag stellt einen Kompressor für solche Bitströme auf der Basis eines LZ77-Wörterbuchansatzes vor. Im Gegensatz zu klassischen Implementierungen werden Wiederholungen auf Bitebene detektiert. Bereits mit kleinen Wörterbuchlängen wird eine effiziente Kompression erzielt. Dies ermöglicht die Implementierung z.B. in FPGAs. Es wird gezeigt, dass der Kompressor mit der Geschwindigkeit praktisch relevanter Tracedatenquellen arbeiten kann.

Schlüsselwörter: Instruktionstrace, On-Chip-Tracing, Kompression, Ausführungsbits, LZ77, FPGA

1 Einleitung

Ein wesentliches Kennzeichen von in eingebetteten Systemen laufender Software ist deren stetig wachsende Komplexität. Damit wird auch der Software-Test zu einer zunehmend kritischen und zeitaufwendigen Aufgabe. Softwaredebugging und Validierung haben einen großen und weiter steigenden Anteil am gesamten Entwurfsaufwand. Das nicht-invasive Beobachten und Aufzeichnen von Systemzuständen, während das System in Echtzeit und in realer Interaktion mit der Systemumgebung läuft, ist ein wesentlicher Bestandteil des Software-Tests. Dieses passive Abtasten wird als *Tracing* bezeichnet. Typische mittels Tracing realisierte Funktionen und Aufgaben sind die nicht-invasive Analyse des Laufzeitverhaltens, die Fehlerdetektion und die Messung der Code-Abdeckung. Für die Zertifizierung sicherheitskritischer Systeme existieren beispielsweise Forderungen nach Echtzeit-Tracedaten für alle Verzweigungsmöglichkeiten im Code.

[1] Brandenburgische Technische Universität Cottbus-Senftenberg, Fakultät für Ingenieurwissenschaften und Informatik, Institut für Medizintechnologie, Großenhainer Str. 57, 01968 Senftenberg, E-Mail: irrgang@b-tu.de
[2] Technische Universität Dresden, Fakultät Informatik, Institut für Technisch Informatik, Nöthnitzer Straße 46, 01187 Dresden, E-Mail: {thomas.preusser,rainer.spallek}@tu-dresden.de

Aus der Integration von Systemen in Systems-on-a-Chip (SoCs) ergeben sich neue Herausforderungen und Problemstellungen für Test und Debugging. Das Hauptproblem ist die erheblich eingeschränkte Beobachtbarkeit und Beeinflussbarkeit der eingebetteten Prozessorkerne. Um zumindest teilweise das Niveau von Logikanalysatoren (LA) und In-Circuit-Emulatoren (ICE) zu erreichen, ist die Integration von dedizierter Trace-Hardware in die SoCs unverzichtbar. Der Hardwareaufwand für ein solches On-Chip-Tracing muss so klein wie möglich sein, um die Kosten für zusätzliche Chipfläche zu minimieren.

Eine wesentliche Herausforderung beim On-Chip-Tracing ist das große Volumen der anfallenden Tracedaten. Die Bandbreite des Traceinterfaces und die Größe des Tracepuffers auf dem Chip sind die hauptsächlichen limitierenden Faktoren. Eine signifikante und unmittelbare Volumenreduktion möglichst nahe der Entstehungsquelle ist beim On-Chip-Tracing unverzichtbar. Die beiden Hauptstrategien sind (a) die Relevanzfilterung durch einen Hardware-Tracemonitor, dessen Triggerbedingungen entsprechend der Analyseaufgabe konfiguriert werden und (b) die Kompression der Tracedaten. Die erste Strategie erfordert vertiefte Kenntnisse des Triggersystems und der Analyseaufgaben. Der Anwender muss entsprechend komplexe Konfigurationen ausführen, um die interessanten Intervalle zu erfassen, ohne dabei relevante Daten zu verlieren. Die Kompression hingegen ist ein Ansatz, der keine Benutzerinteraktion erfordert und nicht auf bestimmte Anwendungen zugeschnitten ist. Sie ist damit ein wertvoller generischer Ansatz, der auch im Anschluss an eine vorgelagerte Relevanzfilterung gewinnbringend eingesetzt werden kann.

In einem SoC können prinzipiell alle Komponenten Quellen von Tracedaten sein. In Bezug auf das Tracing von Prozessoren ist der Instruktionstrace der wichtigste. Eine Vielzahl von Analyseaufgaben kann bereits durch dessen alleinige Anwendung ausgeführt werden. Andere Tracedaten, wie beispielsweise ein Datentrace, sind oft nur im Kontext des aus dem Instruktionstrace rekonstruierten Instruktionsflusses bewertbar.

Bei älteren oder weniger leistungsfähigen Befehlssätzen können nur Verzweigungsbefehle bedingt sein. Der Instruktionsfluss kann sich nur an deren Positionen ändern. Alle dazwischen liegenden Befehle werden streng sequentiell ausgeführt. Für die vollständige, auf dem Binärcode basierende Rekonstruktion des Instruktionsflusses genügt also das Tracing der Verzweigungen. Moderne Befehlssätze hingegen erweitern die bedingte Befehlsausführung auf deutlich mehr oder sogar alle Befehlsgruppen. Ein vollständiger Instruktionstrace muss dann die Ausführungsinformation jedes bedingten Befehls enthalten. Der einfachste Ansatz ist das Aufzeichnen von Ausführungsbits, die anzeigen, ob der jeweilige Befehl ausgeführt wurde. Im Fokus dieses Beitrages steht die komprimierte Kodierung von Strömen aus Ausführungsbits. Ein vollständiger Instruktionstrace muss außerdem die nicht aus dem Binärcode ableitbaren Zieladressen indirekter Verzweigungen mitprotokollieren.

Dieser Beitrag ist wie folgt strukturiert. In Abschnitt 2 werden bestehende Architekturen zur Echtzeit-Kompression von Tracedaten sowie allgemeine Hardwareimplementierungen von LZ-Kompressoren vorgestellt. Die vorgeschlagene Kompressorarchitektur mit Diskussion von Implementierungsaspekten wird in Abschnitt 3 erörtert. Das Format der komprimierten Daten wird in Abschnitt 4 definiert. Nach der abschließenden Bewertung des Kompressors in Abschnitt 5 wird der Beitrag in Abschnitt 6 zusammengefasst.

2 Stand der Technik

Für die Kompression von Zieladressen ausgeführter Verzweigungsbefehle existieren mehrere Lösungen, von denen einige sehr hohe Kompressionsraten erzielen können. Uzelac et al. [UM09] nutzen das Double-Move-to-Front-Verfahren (DMTF) zur Detektion von sich wiederholenden Instruktionsströmen. Die Signalisierung von Wiederholungen erfolgt im Idealfall mit nur einem Bit. Jeder Instruktionsstrom besteht aus der Startadresse und der Länge. Milenkovic et al. [MUMB11] nutzen Caches und Prädiktoren anstatt History-Tabellen zur Detektion von Wiederholungen. Kao et al. [KHH07] schlagen einen LZ-basierten Kompressor für Basisblöcke vor. Die Basisblöcke werden hierbei zur Reduktion der Bitbreite und damit des Hardwareaufwandes des Wörterbuches in Slices aufgeteilt. Basisblöcke und Instruktionsströme sind prinzipiell gleich. Die Länge eines Basisblockes ist die arithmetische Differenz aus der Startadresse des nächsten Basisblockes und der eigenen Startadresse und die Länge eines Instruktionsstromes ist gleich der Anzahl der enthaltenen Instruktionen.

Im Gegensatz zur Kompression von Adressen, Instruktionsströmen und Basisblöcken ist die Kompression von Strömen aus Ausführungsbits ein weitestgehend unbehandeltes Feld. Alle vorstehend aufgeführten Lösungen arbeiten ausschließlich mit Adressen. Nur sehr wenige On-Chip-Trace-Architekturen beziehen die Ausführungsbits überhaupt in das Tracing ein. Im History-Mode von Nexus [IEE12] werden die Ausführungsbits in ein Schieberegister eingetaktet und unkomprimiert in paralleler Form ausgegeben. Die Embedded Trace Macrocell (ETM) [ARM11] als Komponente von ARM CoreSight bildet die Ausführungsbits in einem Bytestrom ab. Hierbei existieren zwei Typen von Einzelbytes, die als P-Header bezeichneten werden. Ein P-Header im Format 2 enthält genau zwei unkomprimierte Ausführungsbits. In einem P-Header im Format 1 sind maximal 16 Ausführungsbits in zwei Binärfeldern kodiert: ein 4-Bit-Zähler repräsentiert null bis 15 konsekutive Befehle, deren Ausführungsbedingung *wahr* war, und ein 1-Bit-Flag zeigt an, ob diesen ein oder kein Befehl mit der Ausführungsbedingung *falsch* folgte.

Die größten Anteile am Volumen von Instruktionstracedaten moderner Prozessoren und Befehlssatzarchitekturen haben (a) Zieladressen ausgeführter Verzweigungen und (b) Ströme aus Ausführungsbits. Zieladressen von direkten Sprüngen können aus dem Binärcode abgelesen werden, so dass deren Aufzeichnung verzichtbar ist. Die Ausführung bedingter direkter Sprünge wird gemeinsam mit allen anderen bedingten Befehlen in den Bitsequenzen aufgezeichnet. Unbedingte direkte Sprünge werden wie sequentielle Befehle behandelt. Mit diesen Festlegungen liegt das Rohdatenvolumen der Ausführungsbits im Median um rund den Faktor 14 über dem der Adressen. Die Streuung über die Programme ist sehr groß. Das Maximum dieses Faktors beträgt rund 22000. Andererseits existieren aber auch Programme, bei denen das Volumen des Adressanteils um den Faktor 128 über dem der Ausführungsbits liegt.

Damit ein Prozessor die benötigten Rohdaten, die Zieladressen und die Ausführungsbits, der Trace-Hardware bereitstellen kann, sollte vorzugsweise dessen Befehlsdecoder so erweitert werden, dass alle tatsächlich zur Ausführung gelangten Befehle diese Daten mit dem Befehlstakt signalisieren.

Das wörterbuchbasierte verlustlose Datenkompressionsverfahren LZ77 ist in vielen verschiedenen Anwendungsgebieten verbreitet. Es ist Bestandteil von DEFLATE, welches ursprünglich von Phil Katz [Deu96b] entwickelt wurde. DEFLATE kombiniert LZ77 mit einer nachfolgenden HUFFMAN-Codierung. Dieser Ansatz wurde auch von Gailly und Adler für deren Dateikompressor GZIP [Deu96c] und für die ZLIB Bibliothek [Deu96a] adaptiert. Neben diesen weit verbreiteten Standardwerkzeugen sind DEFLATE und dessen Derivate allgegenwärtig in der Internetkommunikation bei der verlustlosen Kompression von Grafiken im PNG-Format, des HTTP-Datentransfers oder von SSH-Verbindungen.

Für die Implementierung von LZ77-basierten Kompressoren direkt in Hardware existieren mehrere Vorschläge. Grajeda et al. [GUCP06] stellen einen Hardwarebeschleuniger für die LZ77-Kompression vor, der an einen LEON2-Prozessor angebunden ist. Der Beschleuniger kann zusätzlich eine Burrows-Wheeler-Transformation zur Vorverarbeitung der Daten ausführen. Angewendet wird diese Architektur für eine unmittelbare Online-Datenkompression in vernetzten Kommunikationsgeräten. Rigler et al. [RBK07] beschreiben ein FPGA-basiertes System, bei dem eine LZ77-Kompression und eine HUFFMAN-Codierung in Hardware implementiert sind. Damit sind wesentliche Komponenten für die Hardware-Implementierung von GZIP geschaffen. Ältere Vorschläge für LZ-Hardware-Kompressoren stammen von Chen und Wei [CW99] sowie von Huang et al. [HSM00]. Lin beschreibt eine Hardware-Architektur des LZW-Algorithmus [Lin00]. Alle diese Arbeiten zielen nicht speziell auf die Kompression von Tracedaten ab. Die Arbeit von Huang et al. bildet jedoch den Ausgangspunkt für den Basisblock-Kompressor von Kao et al.

Sowohl Software- als auch Hardwareimplementierungen von LZ77 gehen typischerweise davon aus, auf einem Alphabet aus 256 Symbolen zu arbeiten, die durch 8-Bit-Worte darstellbar sind. Das ist eine sinnvolle Wahl für Daten, die in mindestens ein Byte großen Feldern organisiert sind. Für die überwiegende Mehrzahl der Anwendungsszenarien trifft dies zu, es gibt jedoch auch Ausnahmen. Wolff und Papachristou [WP02] schlagen die Anwendung der LZ77-Kompression auf Bitstrings vor, welche die Testmuster für das automatisierte Testen von VLSI-Chips enthalten. Für diesen speziellen Anwendungsfall schlagen sie weiterhin vor, den hohen Anteil von *Don't Cares* auszunutzen, um die erzielbare Kompressionsrate weiter zu steigern. Die Kompression wird offline ausgeführt und erfordert zwei Läufe über die Daten. Der erste Durchlauf ersetzt die *Don't Cares* mit festen Werten dergestalt, dass bei der Kompression möglichst lange Bitketten im Wörterbuch gefunden werden. Im zweiten Durchlauf erfolgt die eigentliche Kompression. Nach der Übertragung der komprimierten Testmuster in den Chip werden diese in der Hardware dekomprimiert.

Zusammenfassend ist festzustellen, dass keine oder nur sehr einfache Ansätze zur On-Chip-Kompression von Ausführungsbits in Echtzeit existieren. Der Einsatz von Adress- oder anderen Kompressoren, die mit Bytes oder noch breiteren Datenworten arbeiten, ist hierfür nicht praktikabel. Benachbarte Wiederholungen innerhalb eines Bitstromes sind auf der Ebene von Bytes oder höher in der Regel nicht mehr erkennbar. Eine effektive Kompression würde demnach ein größeres Wörterbuch erfordern oder könnte so nicht realisierbar sein. Die Implementierung eines LZ77-Kompressors, der auf Bitebene in Echtzeit arbeitet, scheint ein vielversprechender Ansatz für die Ausnutzung des bisher unausgeschöpften Kompressionspotenzials zu sein.

3 Architektur des Kompressors

Grundsätzlich wäre es denkbar, dass die Ausführung jedes einzelnen Befehls bedingt ist und damit auch ein Ausführungsbit generiert. In diesem ärgsten Grenzfall müsste der Kompressor in jedem Befehlszyklus ein Bit verarbeiten. Praktisch beobachtet werden deutlich geringere Anteile an bedingten Instruktionen. Bei den untersuchten EEMBC-Benchmarks lag deren maximaler Anteil bei 45,8%. Dies erfordert immer noch die Fähigkeit, alle zwei Befehlszyklen ein Bit zu verarbeiten. Der durchschnittliche Anteil bedingter Befehle über alle Benchmark liegt jedoch mit 14,5% noch einmal deutlich unter diesem Wert.

Um einen uneingeschränkten praktischen Einsatz mit einem möglichst vollständigen Trace zu gewährleisten, sollte ein Kompressor, der in jedem Takt ein Ausführungsbit verarbeiten kann, zumindest mit der halben Taktfrequenz des beobachteten Prozessors laufen. Lokale Bursts werden mit einem FIFO von mindestens 10 Bit vollständig ausgeglichen. Taktfrequenzen um 50 MHz werden also in der Regel genügen, wenn die Trace-Quelle ein auf demselben FPGA integrierter Soft-Core-Prozessor ist. Deutlich schneller taktende festverdrahtete externe oder integrierte Prozessoren verlangen schon nach 200 MHz oder mehr. Letztendlich könnte auch ein festverdrahteter Kompressor notwendig sein, um Kerne mit Taktfrequenzen im GHz-Bereich zu bedienen.

Die erzielbare Taktfrequenz hängt stark von der Größe des verwendeten Wörterbuches ab. Ein großes Wörterbuch kommt dem erreichbaren Kompressionsgrad dadurch zugute, dass das Auffinden von Wiederholungen in einem größeren Korpus wahrscheinlicher ist. Die Notwendigkeit der unmittelbaren weiteren Verarbeitung der Eingabedaten schränkt jedoch die Komplexität der realisierbaren Suche ein. Während eine Hardware-Implementierung durchaus einen parallelen Vergleich über alle Wörterbuchpositionen erlaubt, muss aber immer noch die Existenz eines Treffers über eine Berechnung mit einem großen Fanin bestimmt werden. Diese dominiert schnell den kritischen Pfad der Gesamtimplementierung.

Ein weiterer Grund das Wörterbuch eher klein zu halten, liegt im Ressourcenbedarf des *History Buffers*. Während der Ressourcenbedarf der Steuerlogik weitestgehend unabhängig von der Auslegung des Kompressors ist, wächst beim *History Buffer* der Bedarf an Registern und Vergleicherlogik im Gleichschritt mit seiner Größe. Das Ausweichen auf kompaktere RAM-Strukturen ist nicht möglich, da diese den für die parallelen Vergleiche notwendigen gleichzeitigen Zugriff auf den Speicherinhalt nicht erlauben. Ein alternativer Ansatz über eine Hash-Tabelle würde zwar recht schnell die Position eines möglichen Treffers identifizieren, dessen Verifikation müsste daraufhin jedoch durch einen iterativen Vergleich von Speicherbereichen erfolgen. Der Implementierung des *History Buffers* in Registern, die eng mit den parallelen Vergleichern gekoppelt sind, wurde deshalb der Vorzug gegeben.

Bei einer Implementierung in einem FPGA konkurriert die Verwendung von allgemeinen Registern stark mit dem Ressourcenbedarf des überwachten Prozessorsystems. Eine möglichst geringe Beschränkung des Cores und seiner Erweiterungen verlangt nach einem kleinen *History Buffer*. Auch bei einer Realisierung innerhalb der Trace-Architektur eines hartverdrahteten Prozessors beansprucht der *History Buffer* wertvolle Chipfläche, die mit seiner Größe wächst und im nicht überwachten Normalbetrieb ungenutzt bleibt.

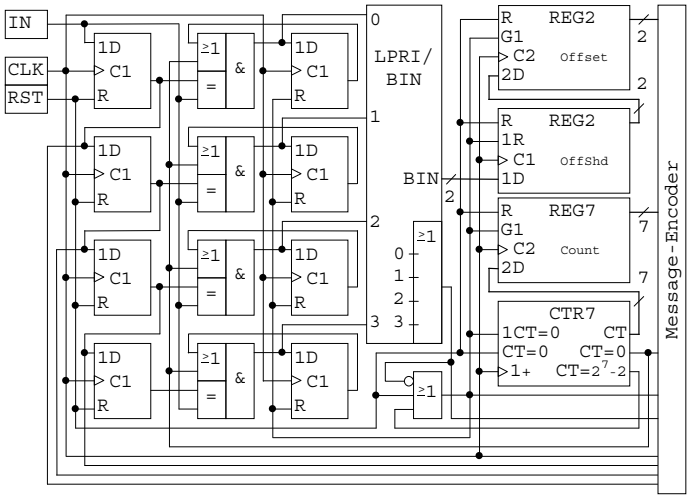


Abbildung 1: Vereinfachtes Schaltbild des Kompressors für Ausführungsbits

Abb. 1 zeigt den Aufbau des Kompressors. Die linke Registerspalte realisiert den *History Buffer*, dessen Inhalt das verfügbare Wörterbuch bereitstellt, als Schieberegister. Die dazu parallele Registerspalte speichert den aktuellen Vergleichszustand für jede Wörterbuchposition. Anfangs zeigen alle Vergleicher eine Übereinstimmung mit dem leeren Wort an. Jede weitere Eingabe eines Ausführungsbits wird mit dem Wert des aktuell im *History Buffer* befindlichen Bits an dieser Stelle verglichen und setzt die Übereinstimmung so fort oder beendet sie. Der Vergleichszustand zeigt also an, ob die seit dem letzten Start eingegebene Bitsequenz mit der an einer Position vorbeigeschobenen übereinstimmt. Beendet eine Eingabe die letzte verbliebene und damit längste Übereinstimmung oder endet die Eingabe gänzlich, so wird diese Übereinstimmung unter Angabe von Position und Länge ausgegeben. Ist sie kurz, so erfolgt die Ausgabe jedoch effizienter als Literal, also in Form einer Kopie der zuletzt konsumierten Bitfolge.

4 Komprimiertes Datenstromformat

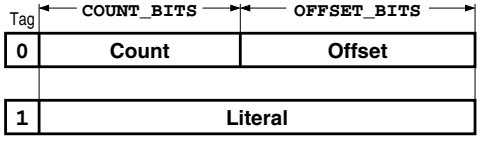


Abbildung 2: Format eines Eintrags im komprimierten Datenstrom

Der komprimierte Datenstrom wird aus Einträgen fester Länge gebildet, deren Format Abb. 2 zeigt. Es gibt zwei Typen von Einträgen: (a) Literale und (b) Rückverweise zur Kodierung von Wiederholungen. Das höchstwertige Bit eines Eintrages, sein *Tag*, differenziert diese beiden Typen. Die Bitlänge eines einzelnen Eintrages wird durch die beiden Codeparameter `COUNT_BITS` und `OFFSET_BITS` bestimmt.

Ein gesetztes Tag-Bit identifiziert ein Literal. Dessen restliche Bits sind zur Dekompression von der höchstwertigen bis zur niederwertigsten Stelle in unveränderter Weise in die Ausgabe zu kopieren. Ein Literal kodiert also eine *unkomprimierte* Bitsequenz der Länge COUNT_BITS+OFFSET_BITS.

Ein gelöschtes Tag-Bit identifiziert (in der Regel) einen Wiederholungseintrag. Das Feld Offset gibt an, an welcher Position im *History Buffer* das jetzt wiederkehrende Muster bereits zu finden ist, wobei dem zuletzt ausgegebenen Bit Null und älteren Ausgaben entsprechend höhere Offsets zugewiesen sind. Die Wahl des Codeparameters OFFSET_BITS bestimmt den Wertebereich dieser Positionsangabe und damit die Größe des Wörterbuches, das der Kompression und Dekompression zugrunde liegt. Die Länge der Wiederholung wird durch das Feld Count spezifiziert. Nicht unterstützt werden hierbei Wiederholungen von weniger als COUNT_BITS+OFFSET_BITS Bits, da die betroffenen Bitsequenzen effizienter in einem Literal kodiert werden können. Um keinen Coderaum zu verschenken, ist die tatsächliche Wiederholungslänge wie folgt zu berechnen:

$$\texttt{COUNT\_BITS} + \texttt{OFFSET\_BITS} + \text{Count}$$

Typische LZ77-basierte Formate würden nach einem Wiederholungseintrag das nächste, nicht mehr übereinstimmende Zeichen implizit als eine Art Literal kodieren. Bei der Verarbeitung von Bitsequenzen kann der Wert dieses ersten nicht mehr übereinstimmenden Zeichens jedoch als Negation des im Wörterbuch auf die Wiederholung folgenden Bits inferiert werden, so dass dieses Bit gar nicht kodiert werden muss.

Die beiden höchsten möglichen Werte des Count-Feldes werden zur Kodierung von Spezialfällen genutzt. Falls Count $= 2^{\texttt{COUNT\_BITS}} - 2$, so entfällt das implizit folgende nicht mehr übereinstimmende Bit. Dies erlaubt die stückweise Kodierung einer Wiederholung, deren Länge die vorgesehene Kapazität des Count-Feldes sprengen würde.

Der Wert Count $= 2^{\texttt{COUNT\_BITS}} - 1$ markiert schließlich das Ende des Bitstromes. Das Offset-Feld wird in diesem Falle genutzt, um die Werte der letzten Bits zu korrigieren. Ist der Offset 0 oder 1, so gibt sein niederwertigstes Bit den tatsächlichen Wert des letzten Bits an. Solche Endmarkierungen finden sich hinter einem abschließenden Wiederholungseintrag, der ja nicht notwendigerweise durch eine fehlende Übereinstimmung beendet wurde, sondern möglicherweise schlicht durch das Ende der Eingabe. Andere Werte im Offset-Feld kodieren die Anzahl der Bits, um die das vorangegangene letzte Literal zu lang war, im Einerkomplement. Diese Korrektur ist notwendig, da es ansonsten keine Möglichkeit gibt, Literale kürzer als $\texttt{COUNT\_BITS} + \texttt{OFFSET\_BITS}$ darzustellen. Die relevanten Bits des letzten Literals befinden sich in den niederwertigen Bits, also rechts ausgerichtet.

Aus der beschriebenen Kodierung ergeben sich einige Randbedingungen für die Codeparameter. Um überhaupt eine Kompression zu ermöglichen, muss $\texttt{COUNT\_BITS} > 1$ sein. Für den Fall $\texttt{COUNT\_BITS} = 1$ könnten nur die beiden Spezialfälle, die Endmarkierung und der Überlauf von Count ohne implizit folgendes nicht mehr übereinstimmendes Bit, kodiert werden. Beide Typen von Einträgen würden so zur Kodierung von $\texttt{COUNT\_BITS} + \texttt{OFFSET\_BITS}$ Bits sogar ein zusätzliches Bit benötigen.

In einer tatsächlichen Realisierung sollte die Wahl von COUNT_BITS die Kodierung typischer Wiederholungslängen ermöglichen. Nimmt man die Implementierung von DEFLATE als vagen Anhaltspunkt [Deu96b, § 3.2.5], so erscheinen Feldgrößen von etwa 8 Bit sinnvoll. Intuitiv sollte auch die Länge des verfügbaren Wörterbuches nicht kürzer als diese darstellbare Länge sein, so dass $\texttt{COUNT\_BITS} \leq \texttt{OFFSET\_BITS}$. Dazu ist jedoch anzumerken, dass das Format (und der implementierte Kompressor) die Kodierung von in sich selbst wiederholenden Bitsequenzen erlaubt. Bei diesen ist die Wiederholungslänge größer als das Offset in das Wörterbuch, so dass die Wiederholung in einen Bereich des Wörterbuches hineinreicht, der erst durch diese Wiederholung definiert wird. Tatsächlich wird ein anfänglicher nur aus Nullen bestehender Wörterbuchinhalt angenommen[1], so dass ein mit n Nullen beginnender Bitstrom als erstes eine Wiederholungsnachricht der Länge n an Offset 0 produziert.

Um die korrekte Kodierung aller Endmarkierung zu ermöglichen, muss das Offset-Feld jede mögliche Zahl von überflüssigen Bits im letzten Literal darstellen können, ohne dass die besonderen Offset-Werte 0 und 1 zur Anwendung kommen. Das heißt:

$$(\texttt{COUNT\_BITS} + \texttt{OFFSET\_BITS} - 1) + 2 \leq 2^{\texttt{OFFSET\_BITS}}$$

$$\texttt{COUNT\_BITS} + \texttt{OFFSET\_BITS} < 2^{\texttt{OFFSET\_BITS}}$$

Die Wahl der Codeparameter bestimmt feste Schranken für die erreichbare Kompressionsrate. Eine stark zufällige Eingabe, in der keine Wiederholung im Bereich des verfügbaren Wörterbuches genutzt werden können, würde als reine Folge aus Literalen kodiert. Dies würde das Datenvolumen um den folgenden Faktor aufblähen:

$$\frac{\texttt{COUNT\_BITS} + \texttt{OFFSET\_BITS} + 1}{\texttt{COUNT\_BITS} + \texttt{OFFSET\_BITS}}$$

Die beste Kompression wird mit Wiederholungseinträgen maximaler Länge erreicht. Solche Einträge können ein Count-Feld mit dem Wert $2^{\texttt{COUNT\_BITS}} - 3$ haben, so dass auf die Wiederholung ein implizites nicht mehr übereinstimmendes Bit folgt; oder sie haben ein Count-Feld mit dem speziellen Wert $2^{\texttt{COUNT\_BITS}} - 2$, bei dem dieses implizite Bit entfällt. In beiden Fällen ergibt sich die erreichbare Kompressionsrate zu:

$$\frac{\texttt{COUNT\_BITS} + \texttt{OFFSET\_BITS} + 1}{2^{\texttt{COUNT\_BITS}} - 2}$$

5 Ergebnisse

Aufgrund der weiten Verbreitung und Typenvielfalt von ARM-Prozessoren werden die Untersuchungen für aktuelle Cortex-Prozessoren durchgeführt. Die Familie der ARM-Prozessoren unterstützt drei Befehlssätze, die sich bezüglich der bedingten Befehlsausführung unterscheiden. Beim Thumb-Befehlssatz des Cortex-M0 können nur Verzweigungsbefehle bedingt sein. Beim erweiterten Thumb2-Befehlssatz des Cortex-M4 existiert der

[1] Das definierte Vorhalten anderer Bitmuster im initialen Wörterbuch könnte die anfängliche Kompressionsrate verbessern, hat aber keine Auswirkungen auf eine eingelaufene Kompression.

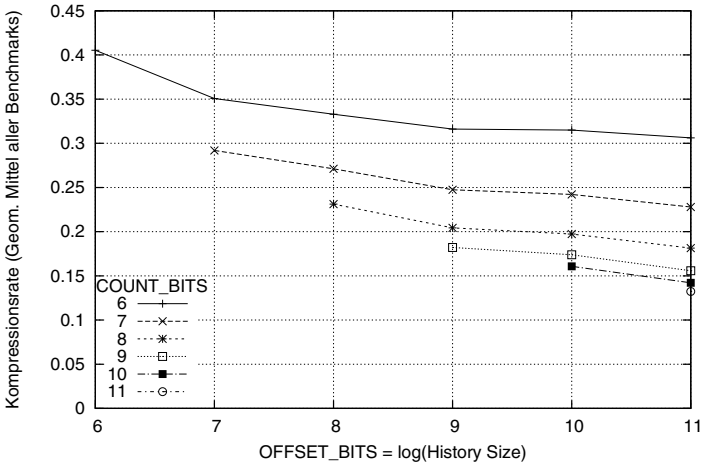


Abbildung 3: Kompressionsrate in Abhängigkeit von den Code-Parametern

Befehl `IT`, der einen If-Then-Else-Block über die unmittelbar nachfolgenden ein bis vier Befehle aufspannt. Die Ausführung dieser Befehle hängt von einer gemeinsamen Bedingung ab, die mit dem `IT`-Befehl angegeben wird und für die mit `if` markierten Befehle gilt. Die Ausführungsbedingung der mit `else` markierten Befehle ist implizit deren Negation. Der ARM-Befehlssatz ist noch flexibler und erlaubt für nahezu jeden Befehl eine eigene, von den anderen unabhängige Ausführungsbedingung. Um sicherzustellen, dass alle Familien und Befehlssätze einbezogen werden, basieren die Untersuchungen auf:

- dem Cortex-M0 und Cortex-M4(F) aus der Familie der Mikrocontroller,
- dem Cortex-A5 und Cortex-A5N aus der Familie der Applikations-Prozessoren und
- dem Cortex-R4F und Cortex-R7F aus der Familie der Realtime-Prozessoren.

Der Einfluss einer zusätzlichen Gleitkommaeinheit ist aus dem Vergleich zwischen dem Cortex-M4F mit FPU und dem Cortex-M4 ohne FPU sowie aus dem Vergleich zwischen dem Cortex-A5N mit der Vektoreinheit NEON und dem Cortex-A5 ersichtlich.

Als einheitliche Basis für die vergleichende Bewertung aller Kompressions-Architekturen und -Methoden ist die EEMBC Automotive Benchmark Software [EEM07] ausgewählt worden. Diese umfasst 16 repräsentative und typischerweise von eingebetteten Systemen ausgeführte Benchmarkprogramme. Zwei Benchmarks sind ausgewiesene Gleitkomma-Benchmarks und weitere zwei enthalten Gleitkommaoperationen. Die Anzahl der jeweils ausgeführten Iterationen übersteigt den Wertevorrat verschiedener Eingangswerte in keinem Fall. Damit ist die Wiedererkennung und Kompression ganzer Iterationen infolge identischer Eingangsdaten ausgeschlossen. Jedes Programm ist mit den Compileroptimierungen *None*, *Size* und *Speed* compiliert worden. Genutzt wurde der Compiler von IAR mit den Optionen *-On*, *-Ohs* und *-Ohz*.

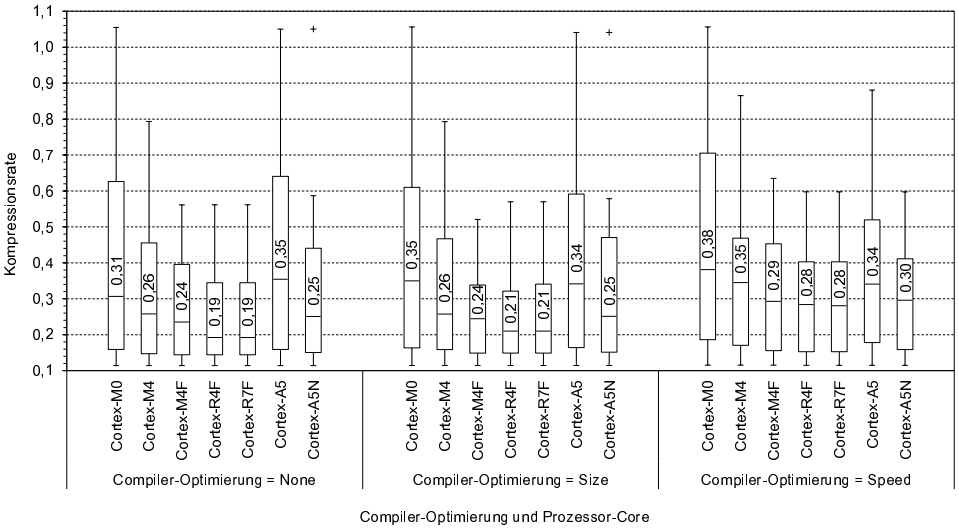


Abbildung 4: Kompressionsrate des Kompressors für Ausführungsbits

Die Abb. 3 zeigt den Einfluss der Code-Parameter auf die erzielbare Kompressionsrate. Innerhalb des betrachteten Parameterraumes erhöhen sowohl eine Vergrößerung von `COUNT_BITS` als auch die von `OFFSET_BITS` die Kompressionsrate. Beide Parameter beeinflussen den Aufwand für die Codierung der Wiederholungseinträge. Der Parameter `OFFSET_BITS` determiniert zudem die Länge des Wörterbuches. Infolge des Codieraufwandes ist ein Absinken der Kompressionsrate mit weiter wachsenden Feldbreiten wahrscheinlich. Alle nachfolgenden Evaluationen basieren auf einer Parametrierung mit einem relativ kleinen Wörterbuch von 256 Bits (`OFFSET_BITS` $= 8$) und einer maximalen Wiederholungslänge von 126 Bits (`COUNT_BITS` $= 7$).

Aus der Abb. 4 sind die damit erzielten Kompressionsraten als Tukey Boxplots ablesbar. Eine solche Darstellung ermöglicht eine detailliertere Sicht auf die Daten und deren Verteilung als die alleinige Angabe eines Mittelwerts. Insbesondere das oft verwendete arithmetische Mittel ist nicht robust genug und daher ungeeignet. Alle Kombinationen aus Prozessoren und Compileroptimierungen sind einbezogen. Somit umfasst jeder Boxplot 16 Benchmarkprogramme.

In der gewählten Konfiguration mit `COUNT_BITS` $= 7$ und `OFFSET_BITS` $= 8$ beträgt die maximal erzielbare Kompression rund 0,113. Das entspricht einer Reduktion des Datenvolumens auf 11,3% der ursprünglichen Größe. In allen Kombinationen wird dieses Maximum von mindestens einem Benchmark nahezu erreicht. Die minimale Kompression beträgt rund 1,067. Das Datenvolumen wird in diesem Fall um 6,7% erhöht. Dieses Minimum tritt nur in seltenen Fällen beim Cortex-M0 und Cortex-A5 auf. In der überwiegenden Mehrzahl von 98% aller Benchmarks wird eine Reduktion des Datenvolumen erreicht. Im Median wird das Volumen um den Faktor drei bis vier reduziert. Für 80% aller Benchmarks wird das Volumen mindestens halbiert. Der Einfluss der Compileropti-

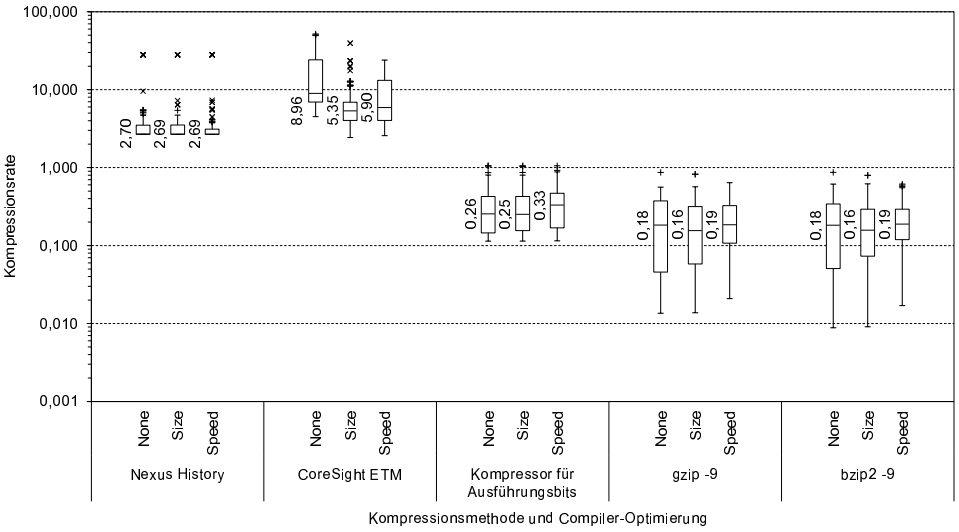


Abbildung 5: Vergleich der Kompressionsraten des Kompressors für Ausführungsbits mit anderen Architekturen und Methoden

mierung ist vergleichsweise klein. Die besten mittleren Kompressionsraten werden bei *None* erzielt. Bei *Size* verschlechtern sich die Ergebnisse nur geringfügig. Für *Speed* ist ein etwas stärkerer Rückgang festzustellen.

Der Prozessorkern hat einen größeren Einfluss auf die Kompressionsrate als die Compileroptimierung. Die schlechtesten Ergebnisse werden mit dem Cortex-M0 erzielt. Allerdings ist dieser auch der einzige Prozessor, bei dem nur Verzweigungsbefehle bedingt sein können. Mit dem Cortex-A5 werden nur geringfügig bessere Ergebnisse erreicht. Durch die Hinzunahme der Vektoreinheit NEON (Cortex-A5N) wird die Kompressionsrate deutlich verbessert. Dieser positive Einfluss einer FPU auf die Kompression ist auch aus dem Vergleich zwischen dem Cortex-M4 und dem Cortex-M4F erkennbar. Die besten Ergebnisse werden mit den Realtime-Prozessoren Cortex-R4F und Cortex-R7F erzielt.

Das Hauptmotiv für die durchgeführten Untersuchungen war die Frage, ob eine nennenswerte Kompression von Ausführungsbits innerhalb des Instruktionstrace von modernen Prozessoren mit moderatem Hardwareaufwand und in Echtzeit möglich ist. Den Ausgangspunkt für die Vergleiche bilden die einzigen beiden On-Chip-Trace-Architekturen, Nexus History und CoreSight ETM, welche die Ausführungsbits überhaupt in den Instruktionstrace einbeziehen. Die Ergebnisse sind in Abb. 5 dargestellt.

Nexus History gibt die Ausführungsbits generell unkomprimiert aus. Aufgrund des Messageformates wird ein Overhead hinzugefügt, so dass sich das Datenvolumen immer erhöht. Im Median wächst das Volumen um den Faktor 2,7. Die von CoreSight ETM erreichten Kompressionsraten sind noch schlechter als die von Nexus History. Relativierend muss hier die abweichende Funktionsweise in Betracht gezogen werden. Die ETM zeichnet nicht nur für jeden bedingten Befehl ein Ausführungsbit auf, sondern auch für jeden un-

Tabelle 1: Ressourcenbedarf des Kompressors für Ausführungsbits

Device	Register	LUTs	Taktrate
XC5VLX50T-FF1136-2 (speed)	550	1635	218 MHz
XC5VLX50T-FF1136-2 (size)	550	1428	107 MHz

bedingten. Ein unbedingter und somit immer ausgeführter Befehl wird in derselben Weise wie ein ausgeführter bedingter Befehl behandelt. Dies führt zu einer Verringerung der Kompressionsrate, wenn auch hier nur die bedingten Befehle als Berechnungsbasis zugrunde gelegt werden. Aber auch dann, wenn dieser Fakt vernachlässigt wird, sind keine wesentlich besseren Resultate als bei Nexus zu erwarten.

Die Vergleiche mit den beiden etablierten On-Chip-Trace-Architekturen verdeutlichen das Kompressionspotenzial in den Ausführungsbits und demonstrieren die Leistungsfähigkeit des hier vorgeschlagenen Kompressors. Gegenüber Nexus History wird das Datenvolumen um einen Faktor von acht oder mehr reduziert.

Abschließend erfolgt der Vergleich der Kompressionsraten mit den für die Kompression von Dateien etablierten Kompressoren bzip2 und gzip. Obwohl der Aufwand für deren vollständige Implementierung in Hardware sehr hoch und oftmals unakzeptabel ist, sind sie dessen ungeachtet eine Referenz für die praktisch erreichbaren Kompressionsraten. Bei beiden Kompressoren wird der Kompressionsaufwand mittels eines Parameters auf der Kommandozeile übergeben. Aufgrund des geringen Einflusses dieses Parameters bei den hier verwendeten Daten ist nur die höchste Aufwandsstufe (`-9`) dargestellt.

Wie erwartet erreichen bzip2 und gzip bessere Kompressionsraten als der hier vorgestellte Kompressor. Diese beiden erzielen ihre Ergebnisse jedoch aufgrund von zusätzlichen Verarbeitungsschritten wie der Huffman-Codierung und Burrows-Wheeler-Transformation sowie aus viel größeren Speicherblöcken für die History. Während gzip mit 32 kB großen Blöcken arbeitet, nutzt bzip2 in Abhängigkeit von der gewählten Aufwandsstufe sogar 100 bis 900 kB große Blöcke. Das liegt um Größenordnungen über der Wörterbuchgröße von 256 Bits (64 Bytes), die in der hier vorgeschlagenen Implementierung verwendet wird. Trotzdem liegt die damit erzielte mittlere Kompression bemerkenswert nahe an denen von bzip2 und gzip. Die Worst-cases sind nur unwesentlich schlechter. Lediglich die Tracedaten, die bereits vom Hardware-Kompressor besonders gut komprimiert werden, werden von bzip2 oder gzip noch einmal deutlich besser verdichtet.

Die Tab. 1 zeigt die Eigenschaften zweier konkreter Hardwareimplementierungen auf einem Virtex-5 FPGA. Für die Ermittlung dieser Werte wurde das Kompressordesign zwischen den Eingangsregistern und den registerbasierten Ausgängen eingebettet, so dass das ISE Synthesewerkzeug die erzielbare Taktfrequenz zuverlässig aus dem kritischen Register-Register-Pfad bestimmen kann. Die angezeigte Ressourcennutzung beinhaltet das komplette Toplevel-Design inklusive der Eingangs- und Ausgangsregister.

Bewertet wurden zwei alternative, funktionell äquivalente Implementierungen. Obwohl beide im generischen, geräte- und herstellerunabhängigen VHDL umgesetzt wurden, ver-

wenden sie verschiedene Ansätze zur Ermittlung des Match-Offsets im Wörterbuch. Diese Berechnung erfordert die Lokalisierung eines gesetzt verbliebenen Match-Signals aus allen 256 Positionen des Wörterbuches. In der geschwindigkeitsoptimierten Variante erzeugt das Synthesewerkzeug einen Reduktionsbaum. Die größenoptimierte Variante nutzt einen Addierer für die Berechnung des Zweierkomplements $-m$ des Match-Vektors m. Aus diesem wird anschließend die bitweise `und`-Verknüpfung `m&-m` gebildet. In deren Ergebnis ist nur das in m äußerst rechts stehende gesetzte Bit gesetzt. Das Umcodieren dieser One-Hot-Codierung in eine binäre Position ist einfacher, so dass insgesamt eine Reduktion um mehr als 200 LUTs erreicht werden kann. Allerdings ist die Carry-Chain des 256-Bit-Addierers bereits recht lang und komplex, so dass die maximale Taktfrequenz erheblich beeinträchtigt wird. Trotzdem ist der Kompressor ausreichend schnell für viele Soft-Core-Prozessoren. Anspruchsvollere Tracequellen wie Hard-Core-Prozessoren erfordern die geschwindigkeitsoptimierte Variante oder gar ein Härten des Kompressors.

Die Ergebnisse zeigen, dass ein auf Bitebene arbeitender LZ77-Kompressor mit einem vergleichsweise geringen Hardwareaufwand eine deutliche Kompression der Tracedaten bewirken kann. Die gewählte Konfiguration mit einer Wörterbuchgröße von 256 Bits ist ein guter Kompromiss zwischen Hardwareaufwand und erzielbarer Kompressionsrate.

6 Zusammenfassung

Tracing als nicht-invasives Echtzeit-Monitoring von Systemzuständen ist für den Software-Test in eingebetteten Systemen unverzichtbar. Die Aufzeichung des Instruktionsflusses ist dabei am wichtigsten. Für komplexe SoCs ist die Integration von Trace-Hardware unentbehrlich. Der Aufwand für ein solches On-Chip-Tracing muss möglichst gering sein. Eine der wesentlichsten Herausforderungen ist die signifikante Reduktion des Tracedatenvolumens. Die Kompression ist eine besonders geeignete Strategie dafür.

Innerhalb des Instruktionstraces von modernen Befehlssatzarchitekturen besteht der größte Teil des Datenvolumens aus zwei verschiedenen Komponenten: Zieladressen indirekter Verzweigungen und Ausführungsbits, die infolge der bedingten Ausführung anderer Befehle entstehen. Während für die Kompression von Adressen mehrere Lösungen existieren, werden Ausführungsbits bisher nur rudimentär oder gar nicht komprimiert. Der vorliegende Beitrag schließt diese Lücke und stellt einen auf Bitebene arbeitenden LZ77-Kompressor für Sequenzen von Ausführungsbits vor. Dessen Implementierung und das Ausgabeformat sind beschrieben worden.

Der vorgestellte Kompressor wurde auf der Basis der EEMBC Automotive Benchmark Suite evaluiert. Jedes der 16 Benchmarkprogramme wurde mit drei Optimierungsleveln compiliert und auf sieben verschiedenen ARM Cortex Prozessoren ausgeführt. Alle drei Familien und alle drei Befehlssätze wurden auf diese Weise in die Untersuchungen einbezogen. Bei 98% aller Benchmarks wird das Datenvolumen reduziert und bei 80% wird es mindestens halbiert. Der Median der Volumenreduktion liegt im Bereich von einem Drittel bis zu einem Viertel. Diese Resultate liegen beachtlich nahe an den von bzip2 und gzip erzielten. Die Implementierung auf einem Virtex-5 FPGA benötigt rund 550 Register, 1500 LUTs und erreicht eine Geschwindigkeit von bis zu 218 MHz.

Literatur

[ARM11] ARM Limited. Embedded trace macrocell architecture specification ETMv1.0 to ETMv3.5, 2011.

[CW99] Jun-Min Chen and Che-Ho Wei. VLSI design for high-speed LZ-based data compression. *IEE Proceedings - Circuits, Devices and Systems*, 146(5):268–278, Oct 1999.

[Deu96a] P. Deutsch. RFC 1950: ZLIB compressed data format specification version 3.3, May 1996.

[Deu96b] P. Deutsch. RFC 1951: DEFLATE compressed data format specification version 1.3, May 1996.

[Deu96c] P. Deutsch. RFC 1952: GZIP file format specification version 4.3, May 1996.

[EEM07] EEMBC. EEMBC autobench 1.1 benchmark software, 2007.

[GUCP06] Virgilio Zuñiga Grajeda, Claudia Feregrino Uribe, and René Cumplido-Parra. Parallel hardware/software architecture for the BWT and LZ77 lossless data compression algorithms. *Computación y Sistemas*, 10(2):172–188, 2006.

[HSM00] Wei-Je Huang, Nirmal Saxena, and Edward J. McCluskey. A reliable LZ data compressor on reconfigurable coprocessors. In *Proceedings of the 2000 IEEE Symposium on Field-Programmable Custom Computing Machines*, FCCM '00, pages 249–258. IEEE Computer Society, Apr 2000.

[IEE12] IEEE-ISTO. The nexus 5001 forum standard for a global embedded processor debug interface version 2.0 (IEEE-ISTO 5001-2003), 2012.

[KHH07] Chung-Fu Kao, Shyh-Ming Huang, and Ing-Jer Huang. A hardware approach to real-time program trace compression for embedded processors. *IEEE Trans. on Circuits and Systems I: Regular Papers*, 54(3):530–543, Mar 2007.

[Lin00] Ming-Bo Lin. A hardware architecture for the LZW compression and decompression algorithms based on parallel dictionaries. *J. VLSI Signal Process. Syst.*, 26(3):369–381, Nov 2000.

[MUMB11] A. Milenkovic, V. Uzelac, M. Milenkovic, and M. Burtscher. Caches and predictors for real-time, unobtrusive, and cost-effective program tracing in embedded systems. *Computers, IEEE Transactions on*, 60(7):992–1005, July 2011.

[RBK07] S. Rigler, W. Bishop, and A. Kennings. FPGA-based lossless data compression using Huffman and LZ77 algorithms. In *Canadian Conference on Electrical and Computer Engineering (CCECE 2007)*, pages 1235–1238, Apr 2007.

[UM09] V. Uzelac and A. Milenkovic. A real-time program trace compressor utilizing double move-to-front method. In *Design Automation Conference, 2009. DAC '09. 46th ACM/IEEE*, pages 738–743, July 2009.

[WP02] F.G. Wolff and C. Papachristou. Multiscan-based test compression and hardware decompression using LZ77. In *International Test Conference*, pages 331–339, 2002.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Test eingebetteter Prozessoren im Zielsystem mit hoher diagnostischer Auflösung

Christian Gleichner, Heinrich T. Vierhaus, Brandenburgische Technische Universität Cottbus-Senftenberg

Abstract: Strukturorientierte Tests können eingesetzt werden, um die Funktionstüchtigkeit hochintegrierter Schaltungen zu überprüfen. Der Produktionstest wird von einer speziellen, schaltungsinternen Testlogik unterstützt, die über dedizierte I/O-Kanäle angesteuert wird. Im Zielsystem ist diese Schnittstelle aber nicht verfügbar. Um einen erweiterten Systemtest mit diagnostischen Fähigkeiten im Feld zu realisieren, bietet es sich an, den Testzugang über vorhandene Standardschnittstellen zu ermöglichen. Die in dieser Arbeit vorgestellte Testschnittstelle erlaubt es, Testroutinen für den im Zielsystem eingebetteten Prozessor durchzuführen, welche über den stets unvollständigen funktionalen Test hinaus auch strukturorientierte Tests mit hoher Fehlerüberdeckung umfassen.

Keywords: Scan-test, Fehlerdiagnose, eingebettetes System, EDT, USB, FlexRay.

1 Einleitung

In vielen technologischen Bereichen nimmt die Komplexität eingebetteter Systeme mit steigenden Ansprüchen bezüglich Qualität und Sicherheit, aber auch aus ökologischen und ökonomischen Aspekten stetig zu. Zum einen wird dies erreicht durch die Anzahl der Prozessoren, Controller und Sensoren und deren Vernetzung, zum anderen durch den rasanten technologischen Fortschritt in der Halbleiterindustrie. Durch die hohe Integrationsdichte und die dadurch erhöhte Sensibilisierung gegenüber potentieller Fehlerquellen während des Produktionsprozesses lassen sich bei der Halbleiterfertigung defekte Chips oder solche mit erhöhter Fehleranfälligkeit nicht vermeiden. Somit werden Ausbeute und Lebensdauer hochintegrierter Schaltungen (IC) reduziert. Da diese vermehrt auch Einzug in sicherheitskritische Anwendungen halten, ist es wichtig, einen hohen Qualitätsstandard und eine hohe Fehler- und Ausfallresistenz zu gewährleisten.

Sei hier das Automobil exemplarisch für ein komplexes elektronisches System betrachtet. Es wäre wünschenswert einen Fehler möglichst frühzeitig zu erkennen, bevor er zu Störung essentieller Funktionen führt. Kommt es zum Teil- oder gar Systemausfall, so ist es darüber hinaus von großer Bedeutung, einen einmal festgestellten Fehler im Nachhinein in der Werkstatt oder als Rückläufer beim Hersteller schnell und eindeutig reproduzieren und diagnostizieren zu können. Die Ursache eines gemeldeten Fehlers in der Fahrzeugelektronik ist aber häufig nicht einwandfrei feststellbar [VS14]. So besteht im Fehlerfall nur die Möglichkeit, Systemkomponenten anhand der Fehlerbeschreibung auf Verdacht auszutauschen. Eine nachträgliche Fehleranalyse durch den Halbleiter-

hersteller erfordert hohen Aufwand, da der IC unter anderem erst von der Platine gelöst werden muss.

Im Produktionstest beim Halbleiterhersteller werden hochauflösende strukturorientierte Verfahren angewandt, um fehlerhafte Chips zu identifizieren und auszusortieren. Hierzu werden in den IC eingebrachte Teststrukturen mit separaten Zugangskanälen genutzt, um eine hohe Fehlerüberdeckung in kurzer Testzeit zu garantieren. Der Testzugang zu dieser Produktionstestlogik steht nach dem Packaging und somit nach dem Aufbringen auf die Platine nicht mehr zu Verfügung.

Das entworfene Konzept realisiert einen für die Diagnose eingebetteter Prozessoren erforderlichen strukturorientierten Test unter Verwendung der schaltungsinternen Produktionstestlogik und serieller Standardschnittstellen. Somit wird ein Test eines ICs mit hoher diagnostischer Auflösung über eine vorhandene Steuergeräteschnittstelle im Zielsystem (Kraftfahrzeug) ohne Demontage des Steuergerätes und der ICs verfügbar gemacht. Dem Steuergerätehersteller kann dadurch ein erweiterter Produktionstest bereitgestellt werden, der weit über den Leiterplattentest hinaus eine nachweisbare hohe Prüfschärfe bietet.

Kann die Störung einer Systemfunktionalität durch die Diagnosemöglichkeit bereits im Feld, d.h. während eines Werkstattaufenthalts, eindeutig auf einen fehlerhaften IC zurückgeführt werden, lassen sich wiederholte Fehlersuchen und teure Reparaturen vermeiden. Der Halbleiterhersteller kann zudem die aus der Diagnose eines defekten ICs gewonnenen Informationen nutzen, um während des Fertigungsprozesses und des Fertigungstests entsprechende Maßnahmen zu ergreifen, mit denen die Chipqualität verbessert werden kann [Ab14].

Um integrierte Schaltungen auch nach dem Fertigungstest noch testen zu können, steht heutzutage bereits die etablierte JTAG-Schnittstelle zur Verfügung [JT01]. Diese ist hauptsächlich für den Boundary-Scan zum Test der Verbindungen zwischen IC und Leiterplatte vorgesehen. Darüber hinaus ist es mittels JTAG möglich, Testvektoren an die primären Eingänge der internen Schaltungslogik anzulegen, um diese erschöpfend, pseudoerschöpfend oder funktional zu testen. Ist der interne Schaltungsaufbau bekannt, kann auch ein strukturorientierter Test, mittels eines SBST-Verfahrens (*Software-based self test*), realisiert werden.

Die neue IJTAG-Schnittstelle soll einen standardisierten und kostengünstigen Zugang zu Testmodulen integrierter Schaltungen auch noch nach dem Produktionstest ermöglichen. Diese internen Testmodule können etwa MBIST-Einheiten zum Test der Speicherbausteine oder LBIST-Einheiten zur Ansteuerung der integrierten Scan-Strukturen sein. Somit macht es IJATG ebenso möglich, die in den IC integrierte Produktionstestlogik für einen diagnostischen Test zu nutzen. Das in dieser Arbeit vorgestellte Konzept stellt eine Alternative zur IJTAG-Lösung dar.

IJTAG nutzt den JTAG-TAP als Zugang zum Gateway-Pfad und zu den daran angeschlossenen Testmodulen [IJ13]. Da der JTAG-TAP und der Scan-Pfad üblicher-

weise mit 5 bis 30 MHz betrieben werden, ist auch die Datenrate für den IJATG-Test entsprechend beschränkt. Da die Testeingaben direkt mit jedem Takt seriell hineingeschoben werden, bedeutet eine JTAG-Clock von 30 MHz eine Datenrate von 30 MBit/s. Der Testzugang über standardisierte Anwendungsschnittstellen kann eine kürzere Testzeit ermöglichen, wenn ein Bussystem mit hoher Datenrate die Grundlage bildet.

Neben dem Testzugang umfasst das entworfene Konzept auch einen integrierten Selbsttest, der strukturelle Fehler des ICs bereits vor einer möglichen Auswirkung auf die Funktionalität des Systems identifizieren kann und somit die Systemzuverlässigkeit erhöht. Um den universellen Einsatz in verschiedenen Anwendungsszenarien zu unterstreichen, wird das entworfene Design im Folgenden als *Universal Scan-Test Interface* (USIF) bezeichnet

2 Universal Scan-Test Interface

Die USIF-Architektur realisiert den Testzugang zum eingebetteten System über HighSpeed-Standardschnittstellen. Über die als Testzugang spezifizierte Schnittstelle werden Testdaten übertragen, in ein definiertes einheitliches Format, welches der zugrundeliegenden Testtechnologie angepasst ist, zwischengespeichert und an den integrierten Test-Controller weitergeleitet. Diese Testdaten können komplette Testmuster für einen direkten Scan-Test, alternativ aber auch hochgradig komprimierte Konfigurationsdaten für einen Selbsttest sein.

Die Testkonfigurationen, wie z.B. der Testtyp, die Testanalyse oder das Ausgabeformat, werden über die im Testsatz enthaltenen Metadaten bestimmt. Testdaten werden entsprechend dem spezifizierten Verfahren in die ICs geladen, um diese strukturell testen zu können. Mittels zusätzlich übertragener Referenzdaten kann eine schaltungsinterne Analyse der Testergebnisse erfolgen. Zudem können die Testantworten auch direkt oder kompaktiert zu Signaturen und ggf. zusätzlich gefiltert über die Standardschnittstelle für anschließende diagnostische Auswertungen ausgelesen werden. Das Konzept wurde bereits in [Gl12, DI13] vorgestellt.

2.1 Aufbau der Testschnittstelle

Die Anbindung an die Scan-Testarchitektur geschieht hauptsächlich über eine Steuerungseinheit und einen oder mehrere Zwischenspeicher für die zu übermittelnden Testdaten. In Abb. 1 ist der konzeptionelle Aufbau der Testschnittstelle dargestellt. Als Zwischenspeicher werden hier separate FIFO-Module verwendet. Der Speicher für die Testeingaben ist das *TestIn*-FIFO. Für die Referenzdaten, die zur Auswertung der Testergebnisse benötigt werden, ist hier das *TestRef*-FIFO vorgesehen. In dem Ausgangsspeicher *TestOut*-FIFO werden die Testantworten abgelegt. Die Testantworten des Scan-Tests können entweder direkt der Scan-Struktur entnommen oder kompaktiert

zu Signaturen von der Testarchitektur abgegriffen werden. Die Analysekomponente setzt die optionale Auswertung on-chip um und kann, im Falle einer anschließenden diagnostischen Auswertung, für die Filterung einer umfangreichen Ausgabe auf fehlerhafte Testantworten sorgen. Um den Testzugang zu steuern, muss eine zusätzliche Controller-Einheit implementiert werden. Diese wertet die empfangenen Metadaten aus, koordiniert somit Schreib- und Lesezugriffe und konfiguriert den Scan-Test.

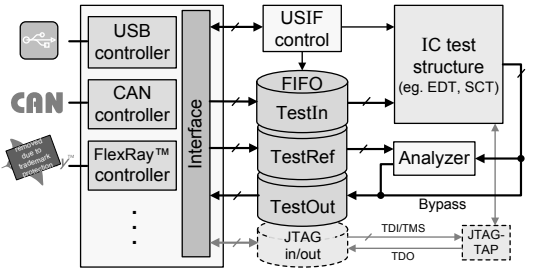


Abb. 1: Aufbau der Testschnittstelle[1]

In der Abb. 1 ist zudem ein optionales JTAG-FIFO zu sehen, das die über eine Standardschnittstelle empfangenen JTAG-Eingangssignale TDI und TMS paketweise zwischenspeichert und für den JTAG-TAP bereitstellen kann. Die Ausgangsdaten aus dem TDO-Signal werden wieder zwischengespeichert, um diese paketweise auslesen zu können. Die Test-Clock TCK und das Reset-Signal TRST werden dabei on-chip erzeugt. Es kann somit auch ein alternativer JTAG-Zugang über eine Standardschnittstelle realisiert werden.

2.2 Testdatenspeicher

Die über die Kommunikationsschnittstelle empfangenen Testeingabedaten werden im *TestIn*-FIFO zwischengespeichert. Das FIFO befindet sich in der USIF-Architektur zwischen dem Empfangspuffer des jeweiligen als Testschnittstelle genutzten Kommunikations-Controllers und der integrierten Testarchitektur (SCT oder EDT). Wie in Abb. 2 dargestellt, wird es mit den aus dem Empfangspuffer entnommenen Daten über den Schreibport der Datenbreite d_{USIF} beschrieben. Diese Datenbreite wird durch die Auslegung der USIF-Architektur bestimmt und wurde für die Umsetzung auf 32 Bit festgelegt. Die Administration der Datenpuffer bei einem Zugriff über die Testschnittstelle wird durch den USIF-Controller übernommen, daher auch durch diesen mit dem Takt der USIF-Clock clk_{USIF} betrieben. Der Lesezugriff erfolgt durch den Test-Controller. Hierbei wird mit Scan-Takt clk_{SC} jeweils ein Datenwort entsprechend der Datenbreite des Eingangsportes i_{SC} der Testarchitektur ausgelesen. In der Abbildung ist zudem das FIFO-Signal *TestIn_locked* als Teil des USIF-Statusregisters zu sehen.

[1] Die Produktionstestlogik wurde anhand des Embedded Deterministic Test (EDT) sowie des an der BTU entworfenen Scan-Controller-based Test (SCT) veranschaulicht.

Hierdurch soll rechtzeitig signalisiert werden, dass das FIFO für den Schreibzugriff über die Testschnittstelle gesperrt ist, um so das System vor einen Deadlock-Zustand zu bewahren. Die weiteren FIFO-Speicher entsprechen in Aufbau und Funktion dem *TestIn*-FIFO, nur das diese sich zwischen den entsprechenden kommunizierenden Komponenten befinden. Das *TestOut*-FIFO, wie in Abb. 2 zu sehen, reicht die Testausgabedaten aus der Test-Architektur, ggf. gefiltert durch den Analyzer, an die Standardschnittstelle weiter.

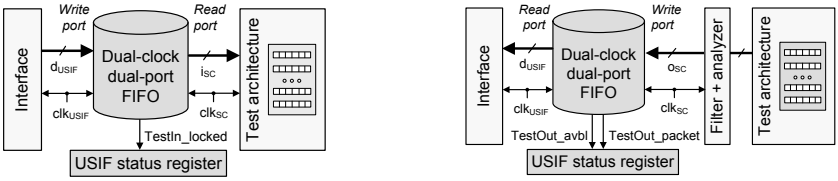


Abb. 2: FIFOs der Testeingabe- (links) und Testausgabedaten (rechts)

2.3 Zentrale Teststeuerung

Der Testzugang und somit der Testablauf wird mittels eines definierten Instruktionssatzes kontrolliert. Die Steuerung des Testzugangs kann durch das Zustandsdiagramm in Abb. 3 nachvollzogen werden. Für den Testzugang durch das USIF wird von einem im speziellen Testmodus befindlichen System ausgegangen. Das heißt, sämtliche von der Standardschnittstelle empfangenen Daten werden als Testdaten, also USIF-Instruktionen bzw. Testmuster für die Testarchitektur, interpretiert.

Vom Ausgangszustand *Idle* wird ein Test durch entsprechende Anweisung gestartet. Hierzu wird der Empfangspuffer des Kommunikations-Controllers ausgelesen, sobald dieser verfügbare Daten enthält. Soll ein interner Selbsttest ausgeführt werden, so wird dieser über den Zustand *RunBIST* initiiert. Im Falle des Zugangs zur internen Testlogik ist zunächst eine Authentifizierung erforderlich. Das in der Abbildung eingerahmte Teildiagramm soll den vor unbefugtem Zugriff geschützten Bereich darstellen.

Im zentralen Zustand *TestControl* findet die Dekodierung der empfangenen USIF-Instruktionswörter statt, um den entsprechenden Zugriff zu initiieren. Um eine Konfiguration, also ein Schreibzugriff auf ein adressiertes Register vorzunehmen, wird der Zustand *WriteConfig* genutzt. Ein Lesezugriff auf ein Konfigurationsregister kann durch den Zustand *ReadConfig* geschehen. Hierbei wird der Registerinhalt angefragt und im darauffolgenden Takt, nachdem der Wert geladen ist, im Zustand *ForwardConfig* in den Sendepuffer des Kommunikations-Controllers geschrieben. Die gestrichelten Pfeile in der Darstellung des Zustandsdiagramms sollen den Zugriff bzw. den Einfluss eines Zustandes auf die entsprechenden Komponenten verdeutlichen.

Ein Lesezugriff auf den Testausgabespeicher wird durch die Zustände *ReadData* und *ForwardData* realisiert. Es wird jeweils ein Datenwort aus dem adressierten

Speicherblock entnommen und an den Sendepuffer des Kommunikations-Controllers weitergereicht. Bei einem Schreibzugriff wird zunächst in den Zustand *WaitData* gewechselt, um auf den Empfang eines Testdatenpaketes mit einem im Instruktionswort spezifizierten Umfang zu warten. Meldet der Kommunikations-Controller verfügbare Daten, wird eine Leseanfrage gestellt, um das erste Datenwort im Zustand *WriteData* in den adressierten Testspeicher zu schreiben. Hier wird in jedem weiteren Takt ein Datenwort gelesen und an den Testspeicher weitergeleitet. Dies wird wiederholt, bis das komplette Datenpaket in den Testspeicher übertragen ist.

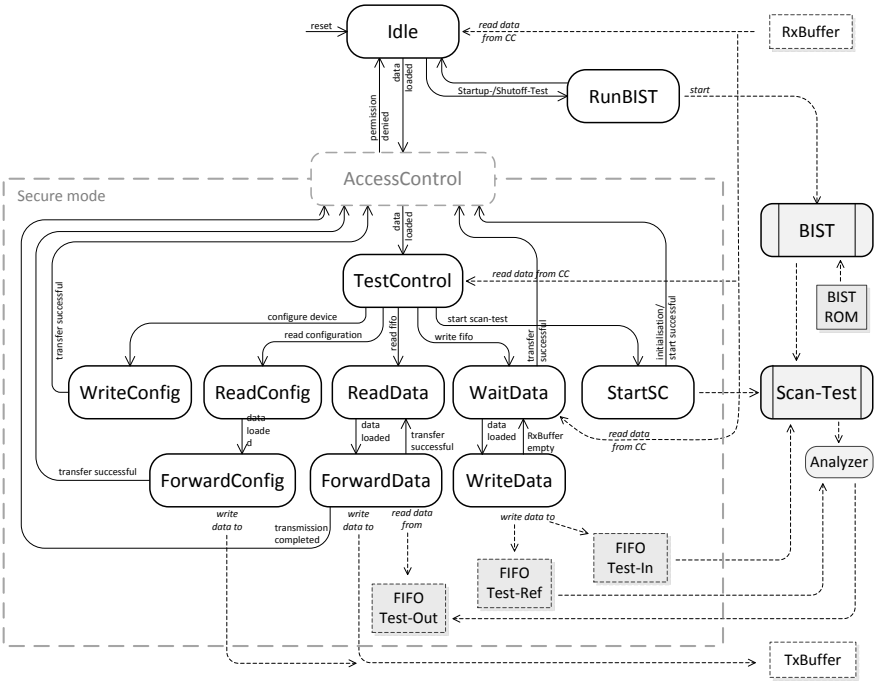


Abb. 3: Vereinfachtes Zustandsdiagramm des USIF-Controllers

Durch den Zustand *StartSC* wird ein Scan-Test initiiert. Hierfür werden die Testkomponenten initialisiert und der Test-Controller gestartet. Nachdem der Scan-Test aktiviert ist, wird direkt wieder in den Zustand *TestControl* gewechselt, um auf den Empfang von Datenpaketen als Testeingaben für den Scan-Test zu warten. Der Test-Controller entnimmt die Testeingaben dem Testspeicher *TestIn*-FIFO.

Die Übermittlung der Testdaten erfolgt paketweise, um den Scan-Test in Zyklen, je nach verfügbaren Testdaten in den Testspeichern, auszuführen. Das heißt, der Testmustersatz wird auf Anwenderseite in Pakete spezifizierter Größe aufgeteilt. Für die eigentliche Übertragung auf dem Buskanal werden die Pakete in Nachrichten dem zugrunde-

liegenden Busprotokoll entsprechend unterteilt. Dieses Schema ist in Abb. 4 veranschaulicht.

Die Übertragung der Testantworten erfolgt auf gleiche Weise. Enthält der Testausgabespeicher ein Datenpaket in entsprechenden Umfang, so wird dieses ausgelesen, um Speicherplatz für weitere Testausgaben freizugeben. Das Datenpaket wird wieder im entsprechenden Nachrichtenformat über den Bus übertragen. Auf Anwenderseite können die Datenpakete entweder direkt ausgewertet oder zu einem kompletten Testantwortsatz zusammengefügt werden.

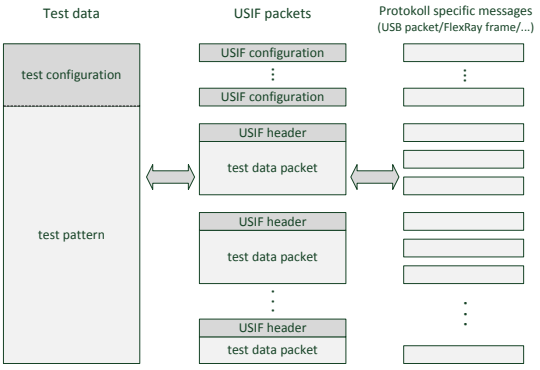


Abb. 4: Paketweiser Datenaustausch

2.4 Test-Architektur

Die industriell genutzten Testtechnologien sind in der Regel Eigentum spezieller EDA-Hersteller, wie z.B. Mentor Graphics, Synopsys oder Cadence, und damit nicht frei verfügbar. Eine der meistgenutzten Methoden ist der *Embedded Deterministic Test* (EDT) von Mentor Graphics [Ra04]. Die Entwicklung der USIF-Architektur erfordert aber eine Lösung, welche frei und ohne Restriktionen verwendet werden kann. Hierzu wird der an der BTU Cottbus entwickelte Scan-Controller als interne Testumgebung in die Architektur eingebunden und somit der sogenannte *Scan-Controller-based Test* (SCT) umgesetzt. Die Basisarchitektur des Scan-Controllers ist in Kooperation mit Infineon Technologies AG im BMBF-Verbundprojekt AZTEKE entstanden [AZ05]. Die Grundlage des Scan-Controllers bildet eine STUMPS-basierte Architektur und beherrscht den strukturorientierten Test sowohl auf statische als auch auf dynamische Fehler. Die Erzeugung der Testmuster beruht dabei auf Pseudozufallssequenzen eines *Linear Feedback Shift Registers* (LFSR) mit nachgeschalteter Musteroptimierung. Die Rückkopplung des LFSR kann innerhalb der Testinitialisierungsphase, aber auch während des Tests, über einen Feedback-Parameter konfiguriert werden. Es handelt sich um ein sogenanntes *Multiple-Polynomial LFSR* (MP-LFSR). Die Komprimierung der Testantworten findet über ein *Multiple Input Signature Register* (MISR) statt. Zusätzlich zu der dadurch umgesetzten *Time Compaction* kann optional eine *Space Compaction*

durch einen XOR-Tree konfiguriert werden. Detaillierte Ausführungen zum Aufbau und Funktionsweise des Scan-Controllers sind in den Vorarbeiten [Ga04, Fr07, Ko08, Ko10] zu finden. In der prototypischen Umsetzung wurde sowohl der SCT als auch der EDT realisiert.

2.5 Testszenarien

Das Ziel dieser Entwicklung ist es, mit derselben Testumgebung optional die folgenden Anwendungsfälle für den Offline-Test zu unterstützen:

- den Produktionstest
  - nach Packaging beim IC-Hersteller,
  - nach Leiterplattenbestückung beim ECU-Hersteller,
- den on-chip Selbsttest im Zielsystem,
- die Fehleranalyse im Zielsystem und
- das Monitoring während des Normalbetriebs.

Diese Szenarien wurden in [Gl12] vorgestellt.

Um die verschiedenen Auswertungsmöglichkeiten umzusetzen, wurden die Analysekomponente sowie ein zusätzlicher BIST-Controller, der den Selbsttest realisiert, entworfen. Dadurch werden folgende Modi der Testanalyse unterstützt:

- *LogAll*, die Ausgabe aller ausgelesenen Testergebnisse,
- *LogFaults*, die Ausgabe der fehlerhaften Testergebnisse inklusive zugehöriger Zeitstempel,
- *PassFail*, die on-chip Auswertung mittels off-chip Referenzdaten, und
- *BIST*, die on-chip Auswertung mittels on-chip Referenzdaten.

3 Applikation

Die Steuerung des USIF-Controllers und somit des diagnostischen Tests soll von einem Anwendersystem aus, das die Kommunikation mit dem peripheren USIF-Gerät beherrscht, möglich sein. Hierzu wurde das Anwendungsprogramm *Universal Scan-Test Interface Applikation*[2] implementiert. Dieses Programm stellt die Schnittstelle der Testumgebung zum Anwender dar. Für die Verbindung zum zu testenden Gerät wird die USB-Schnittstelle[3] genutzt.

[2] Java-Applikation inkl. Benutzeroberfläche, die Eingabemasken zur Konfiguration der Testumgebung und zur Eingabe der Testdaten bereitstellt.
[3] Hierzu wurde als USB-API mit zugehörigem Treiber das Open-Source-Projekt *libusb-win32* eingebunden.

Durch den Anwender angegebene Testdaten[4] können eingelesen, entsprechend umgewandelt und über die zum USIF-Gerät hergestellte Verbindung übertragen werden. Mittels eines Scan-Prozesses ist der Testablauf, die paketweise Übertragung der Testeingaben und -ausgaben über die USB-Schnittstelle und die Testauswertung, automatisiert. Hierzu wird zunächst der Testsatz aus der eingelesenen Datei je nach Format extrahiert und in einem Bytearray für die Testeingaben *arrayIn* und ggf. in einem separaten Array für die Referenzwerte *arrayRef* abgespeichert. Der weitere Ablauf ist in Abb. 5 dargestellt.

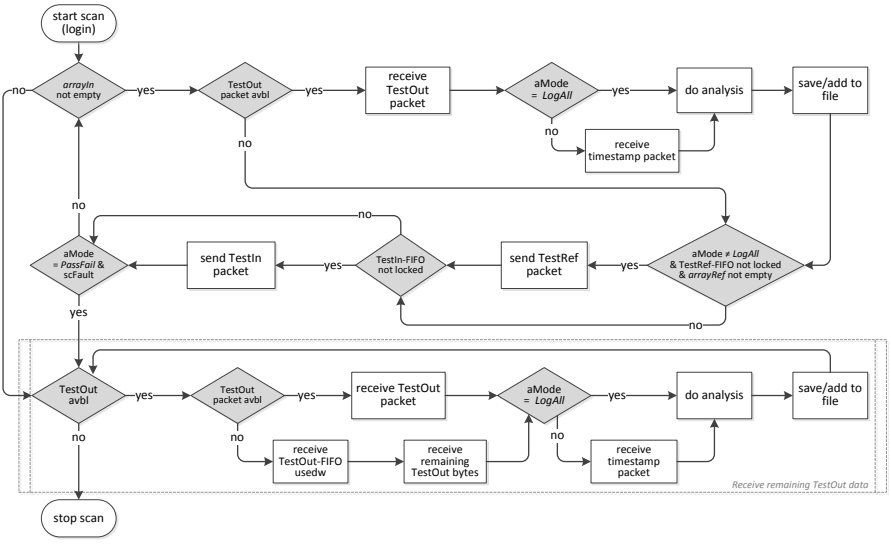


Abb. 5: Datenfluss des Scan-Prozesses

3.1 Extraktion der Testmuster

Das Testprogramm für den Scan-Controller wird durch einen zugehörigen ATPG-Prozess erstellt und als Bitstrom aufbereitet in einer Pattern-Datei abgespeichert. Somit kann die Applikation die erstellte Pattern-Datei einlesen, direkt als Bitstrom interpretieren und an das USIF übertragen.

Die durch kommerzielle ATPG-Programme erzeugten Testmuster liegen in einem standardisierten Format einer *Pattern Description Language* (PDL) vor. Gebräuchliche Pattern-Formate sind *Standard Test Interface Language* (STIL1450, STIL2005), *Core Test Language* (CTL), *Waveform Generation Language* (WGL), *Texas Instruments Test Description Language* (TDL 91). Diese werden üblicherweise durch ATEs interpretiert, um die BIST-Architektur im Fertigungstest anzusteuern.

[4] Pattern-Dateien in SCT-, STIL-, CTL-, TDL-Formaten.

Die Applikation setzt eine Pattern-Extraktion aus einer solchen Dateischnittstelle um. Somit kann ein durch das USIF angesteuerter EDT mittels vorliegender EDT-Pattern betrieben werden. Die USIF-Applikation unterstützt hierzu die Pattern-Formate STIL, CTL und TDL91.

4 Auswertung

4.1 Testeigenschaften

Um Messungen der Testzeiten für verschiedene Schaltungen und Testkonfigurationen vorzunehmen, wurde jeweils die zu testende Schaltung in entsprechender Auslegung der Scan-Struktur und die USIF-Architektur, inklusive der Testarchitektur und der Kommunikations-Controller, synthetisiert und auf FPGA-Basis implementiert. Der Testablauf wird komplett durch die Applikation auf dem PC gesteuert.

Den anschließenden Betrachtungen lag der folgende Versuchsaufbau zugrunde:

- Altera Cyclone II FPGA (EP2C70F896C6N),
- Aufsteckplatine für Schnittstellen:
  - USB Transceiver Texas Instruments TUSB1106, USB 2.0-Buchse Typ Mini-B,
  - FlexRay Transceiver NXP Semiconductors TJA1080A , D-Sub DE9-Buchse,
- Applikation betrieben auf PC Intel Core i7-2600 CPU 3,40 GHz (Arbeitsspeicher 8 GB), OS Windows 7 (64 Bit),
- Testzugang: USB FullSpeed, Bulk-Transfer.

Für die Untersuchung von Testzeiten wurden ISCAS89-, ITC99-Schaltungen und VLIW-Prozessoren[5] in verschiedenen Auslegungen der Scan-Struktur herangezogen. Als Testarchitektur wurde der Scan-Controller eingebunden und dementsprechend die komprimierten Testsätze durch den zugehörigen ATPG-Prozess ermittelt.

Der Ermittlung der Testzeit lag folgende Konfiguration zugrunde:

- Fehlermodell: Stuck-at,
- Scan-Clock: f_{SC} = 20 MHz,
- Scan-Controller Eingangsport: i_{SC} = 16 Bit, Ausgangsport: o_{SC} = 16 Bit,
- Scan-Modus: *Decompress* (LFSR-basierte Testmustererzeugung),
- Kompaktierung: *Mixed* (MISR + XOR-Tree),
- Analysemodus: *LogAll*, Analyse-Rate: (a_{Rate}, a_{Base}) = (m_{Chain}, m_{Chain}-*1*)[6],
- Paketgröße: g_{Packet} = 512 Byte.

[5] Superskalare Architektur mit statischer Programmablaufplanung. Die generische Hardware-Beschreibung ermöglicht es, den VLIW-Prozessor in diversen Konfigurationen zu synthetisieren [Sc06].
[6] Es wird hier jeweils eine Signatur pro Testmuster (bestehend aus m_{Chain} Testvektoren) erzeugt.

In Tab. 1 ist eine Auswahl der untersuchten Schaltungen zusammengefasst. Zu jeder Schaltung der Spalte 1 wurde ein Scan-Design mit unterschiedlicher Anzahl an Scan-Ketten erstellt. In der zweiten Spalte ist die Scan-Struktur dargestellt. Da die Scan-Flipflops zu möglichst gleichlangen Ketten aufgeteilt werden, ergibt sich die Scan-Tiefe m_{chain} aus der spezifizierten Anzahl an Scan-Ketten n_{chain}. Die Spalte 3 enthält die Anzahl der durch den ATPG-Prozess erhaltenen Testmuster. Die durch diese Testmuster erreichbare Fehlerüberdeckung ist in der darauffolgenden Spalte zu finden. Der Datenumfang des komprimierten Testsätze ist in Spalte 5 zu sehen. Als komprimierter Testsatz ist das ermittelte Scan-Controller-Programm zu verstehen [Ga04, Ko08, Gl12]. In der Spalte 6 ist die nötige Testzeit für einen Selbsttest, dem der komprimierte Testsatz für einen deterministischen Test on-chip zur Verfügung steht, dargestellt. In den weiteren Spalten sind die mittels des beschriebenen Versuchsaufbaus eruierten mittleren Testzeiten zu sehen.

	Testschaltung	Scan-Ketten n_{chain}/m_{chain}	Testmuster n_{Pat}	Fehlerüber-deckung (stuck-at) %	Testsatz [kByte]	Zeitaufwand Selbsttest [ms]	Testzeit USB [s]	Testzeit FlexRay[7] [s]
ISCAS89	s13207	128/7	262	94,10	10,03	0,256	0,104	-
		256/4			8,97	0,230	0,089	0,396
	s15850	64/11	155	94,30	9,14	0,234	0,093	-
		128/6			8,07	0,207	0,084	0,310
ITC99	b18	256/11	530	95,38	88,44	2,264	0,906	3,592
		512/6			97,82	2,504	1,029	-
	b19	256/22	538	95,32	176,75	4,525	1,774	8,543
		512/11			193,27	4,958	2,059	-
VLIW16S4[8]		256/12	454	96,21	60,39	1,546	0,541	2,509
		512/6			63,48	1,625	0,620	-
VLIW32S8		256/18	790	98,56	261,62	6,698	2,724	11,067
		512/9			279,63	7,159	2,949	-

Tab. 1: Testaufwand, Fehlerüberdeckung und Testzeit

Für den Selbsttest kann in jedem Scan-Takt ein Datum aus dem komprimierten Testsatz entnommen und dieses an den Eingangsport des Scan-Controllers angelegt werden. Somit ist die Testzeit nur von der Anzahl der Eingaben und der Taktfrequenz des Scan-Controllers f_{SC} abhängig. Für ein Testsatzumfang von g_{Test} Byte und einem Eingangsport der Datenbreite i_{SC} gilt demnach:

[7] Für den Testzugang über FlexRay wurden nur die Testzeiten zu dem jeweils kompakteren Datensatz ermittelt.
[8] Dem Bezeichner ist hier die Anzahl der parallelen Datenpfade (Slots) und die Datenbreite eines Datenpfades zu entnehmen. Der VLIW16S4 bspw. wurde als 16-Bit-Architektur mit 4 Slots synthetisiert.

$$t_{BIST} = \frac{g_{Test}}{i_{SC}/8} * \frac{1}{f_{SC}} \Rightarrow t_{BIST} = \frac{n_{in}}{f_{SC}},$$

wobei $n_{in} = 8g_{Test} / i_{SC}$ die Anzahl der Eingangsdatenworte des Testsatzes ist.

Die Werte für die Testzeit eines über die Standardschnittstellen durchgeführten Scan-Tests wurden durch die Applikation, über die der kompletten Testlauf organisiert wird, ermittelt. Ein Testlauf umfasst hier das Einlesen der Testdaten aus einer spezifizierten Pattern-Datei, der in Abb. 5 beschriebene Scan-Prozess und das Aufbereiten und Speichern der Testergebnisse. Die Testzeit lässt sich folgendermaßen ausdrücken:

$$t_{Test} \approx \frac{g_{Test} + g_{Res}}{r_{Eff}} + t_{Init} + t_{Req} n_{Packet} + t_{Res}.$$

Sämtliche Testdaten, deren Umfang sich sowohl aus dem der Testeingaben g_{Test} als auch aus dem der Testantworten g_{Res} ergibt, werden mit der effektiven Datenrate r_{Eff} übertragen. Hinzu kommen die Zeiten, die für die Initialisierung des Tests t_{Init} und für die Auswertung der Testantworten t_{Res} benötigt werden. Außerdem fällt eine bestimmte Zeit für die Statusanfragen t_{Req} pro Datenpaket an.

Die Initialisierungsphase der Applikation besteht neben der Konfiguration des USIF-Gerätes hauptsächlich aus der Extraktion der Testdaten aus der einzulesenden Pattern-Datei. Die hierfür aufzuwendende Zeit ist also vor allem von dem Datenumfang des Testsatzes abhängig, fällt aber gegenüber der Sendezeit eines umfangreichen Testsatzes relativ gering aus. Da der Analysemodus *LogAll* gewählt und kein Referenzdatensatz spezifiziert wird, fällt auch keine umfangreiche Auswertung der Testergebnisse durch die Applikation an. Somit kann sowohl die Initialisierungszeit als auch die Analysezeit hier vernachlässigt werden. Die Gesamttestzeit wird also auf die Übertragung der Testdaten und der dazu nötigen Kommunikation (Status- u. Request-Botschaften) beschränkt.

Sei hier der Zugang exemplarisch über USB betrachtet, mit den Testzeiten aus Spalte 7. Es ist zu erkennen, dass die Testzeit lediglich von der Größe des Testsatzes abhängt. Der eingestellte Scan-Takt und die Scan-Tiefe, die die Zeit für die Shift-Phase des Scan-Tests bestimmt, wirkt sich nicht auf die Gesamtzeit aus, da zum einen der Test parallel zur Übertragung weiterer Testdaten geschieht. Zum anderen ist die on-chip Verarbeitungsgeschwindigkeit um ein Vielfaches höher als die Übertragungs-geschwindigkeit, sodass die Applikation nicht auf das Hineinschieben der Testmuster in die Teststrukturen warten muss. Da aber vor der Übertragung der Datenpakete sichergestellt werden soll, dass genügend freier Speicherplatz im Empfangspuffer des USB-Controllers bzw. auszulesender Speicher im *TestOut*-FIFO vorhanden ist, erfolgt eine regelmäßige Abfrage des USB-Statusendpunktes. Das bedeutet, es entstehen Wartezyklen pro Datenpaket, die sich negativ auf die Testzeit auswirken.

Da die Übertragung paketweise geschieht und für jedes Paket ein nicht unerheblicher administrativer Aufwand durch die Applikation hinzukommt, ist es hier günstig, die Testzeit über die Zeit pro Paket auszudrücken. Die Paketanzahl n_{Packet} ergibt sich aus dem Umfang des zu sendenden Testsatzes g_{Test}, des zu empfangenen Testantwortsatzes g_{Res} und der spezifizierten Paketgröße g_{Packet}. Da über die Applikation die Gesamtzeit eines Testlaufs ermittelt wurde, kann auf die Zeit pro Paket wie folgt geschlossen werden:

$$t_{Packet} = \frac{t_{Test}}{n_{Packet}} = \frac{t_{Test}}{\lceil g_{Test}/g_{Packet} \rceil + \lceil g_{Res}/g_{Packet} \rceil}.$$

Die Zeiten für die Initialisierung und die Auswertung des Scan-Tests fließen hier in die Paketzeit mit ein. Diese Darstellung der Paketzeit korreliert mit den gemessenen Werten (siehe Tab. 1). Für die unterschiedlichen Testschaltungen ergeben sich in etwa gleiche Zeiten für ein Datenpaket. Die Paketzeit beträgt hier im Mittel um die 5 ms. Daran ist zu sehen, dass ein erheblicher Zeitaufwand durch die Applikation verursacht wird. Über den Bus wird annähernd die volle Datenrate erreicht. Für die USB-Verbindung wurde eine Datenrate für die Nutzdaten von etwa 8 MBit/s ermittelt. Das heißt, ein Paket von 512 Byte kann in 0,512 ms übertragen werden. Der fast zehnfache Wert für ein Paket ist bedingt durch die zeitaufwendigen Statusabfragen.

Die Applikation wurde nicht auf optimale Laufzeit, sondern in erster Linie auf garantierten fehlerfreien Datenaustausch ausgelegt. So wird bspw. vor und nach jedem Zugriff, innerhalb der Sende- und Empfangsfunktion (siehe Abb. 5), auf eine aktualisierte Statusabfrage, die durch den USB-Host mittels Polling im Intervall von 1 ms erfolgt, gewartet. Es wird vorher sichergestellt, dass ein Zugriff erfolgen darf und danach geprüft, ob der Zugriff erfolgreich war.

4.2 Hardware-Aufwand

Um eine Aussage über den Hardware-Aufwand des entwickelten Designs zu machen, wurden durch Synthese, basierend auf der *Si2 NanGate FreePDK45 Generic Open Cell Library*, die Logikzellen und die dadurch eingenommene Chipfläche der einzelnen Komponenten ermittelt. In Tab. 2 sind die resultierenden Chipflächen und der relative Aufwand dargestellt. Hierzu wurde der Hardware-Aufwand exemplarisch einem VLIW-Prozessor, in einer Auslegung als 32-Bit-Architektur mit 8 Slots, gegenübergestellt. Dieser enthält eine Scan-Struktur mit 256 parallelen Scan-Ketten. Dementsprechend wurde der Scan-Controller für 256 Scan-Ketten ausgelegt. Dies betrifft den Eingangsport (für zwei 8-Bit-Adressen) und interne Komponenten wie ALU und MISR.

Die internen Speicherblöcke, wie die Testdatenspeicher des USIF-Controllers und der ROM des Selbsttests, sind hier nicht in die Betrachtung einbezogen worden. Diese Speicher sind in der implementierten Architektur generisch ausgelegt, um die Größe je nach Anwendungsfall anpassen zu können.

Komponenten	Chipfläche in µm²	Relativer Aufwand zur Beispiel-CUT
VLIW32S8	189.993	
Scan-Controller	6.487	3,41 %
USIF-Architektur	3.957	2,08 %
USIF-Controller	2.269	1,19 %
Config-Register	658	0,35 %
BIST-Controller	209	0,11 %
Analyzer	791	0,42 %

Tab. 2: Hardware-Aufwand der USIF-Architektur

Es ist zu erkennen, dass das Design für den Testzugang zur internen Produktionstestlogik gegenüber einem komplexen Prozessor sehr gering ausfällt. Da im eingebetteten System bereits vorhandene Standardschnittstellen und Sicherheitsmodule genutzt werden sollen, wird also nur verhältnismäßig wenig Zusatzlogik für die Verknüpfung der Anwendungsschnittstelle mit der Testarchitektur benötigt.

5 Zusammenfassung

Mittels des entworfenen Konzeptes ist es möglich, eine feingranulare Fehlerdiagnose von hochintegrierten Schaltungen auch in der Postproduktionsphase durchzuführen.

Für eingebettete Prozessoren bedeutet dies, dass im Falle einer Fehlfunktion des Systems eine vorhandene Anwendungsschnittstelle als Testzugang für einen detaillierten diagnostischen Test dient, somit also keine aufwendige Demontage für nähere Untersuchungen notwendig ist. Neben der gewonnenen Diagnosefähigkeit im Feld kann der Testzugang bereits den Produktionstest beim Komponentenhersteller unterstützen.

Eine weitere Testumgebung ist die feingranulare Fehleranalyse von Rückläufern. Der IC kann, ohne diesen von der Platine lösen zu müssen, beim Komponenten- oder Halbleiterhersteller detailliert untersucht werden, um mögliche Fehler in der Schaltungsstruktur zu diagnostizieren.

Ein wichtiger Punkt, der hier nochmals Erwähnung finden muss, ist der Schutz vor unbefugtem Zugriff. Der Zugang über standardisierte Schnittstellen zu internen Strukturen eines Prozessors, erfordert adäquate Maßnahmen gegen unautorisierte Kommunikation und missbräuchliche Anwendungen. Es gibt bewährte industriell genutzte Hardware-Lösungen, wie bspw. der von deutschen Automobilherstellern angewandte SHE-Standard, und darauf basierende oft herstellerspezifische Verfahren, die Authentifizierung und Verschlüsselung für Systemzugriffe gewährleisten [SH09].

Bei einer angemessenen Lösung für sicheren Zugriff bietet die präsentierte Testanbindung, die gegenüber einem komplexen IC eine Zusatzlogik mit relativ geringem Hardware-Aufwand darstellt, einen erheblichen Nutzen bezüglich Systemzuverlässigkeit und Diagnosefähigkeit.

Literaturverzeichnis

[Ab14] Abelein, U.; Cook, A.; Engelke, P.; Glaß, M.; Reimann, F.; Russ, T.; Teich, J.; Wunderlich, H.-J.: Non-Intrusive Integration of Advanced Diagnosis Features in Automotive E/E-Architectures. In Proc. IEEE Design, Automation & Test in Europe Conference & Exhibition, S. 1-6, 2014.

[AZ05] Applikationsspezifische Testmethodik für hochkomplexe Systeme der Kommunikations- und Kraftfahrzeugtechnik: BMBF-Verbundprojekt AZTEKE; Abschlussbericht. Technischer Bericht, ATMEL Germany GmbH, Infineon Technologies AG, Philips Semiconductors GmbH, 2005.

[DI13] Durchgängige Diagnosefähigkeit in Halbleiterbauelementen und übergeordneten Systemen zur Analyse von permanenten und sporadischen Elektronikausfällen im Gesamtsystem Automobil: BMBF-Verbundprojekt DIANA, Technischer Bericht, Audi AG, Conti Temic Microelectronic GmbH, Infineon Technologies AG, Zentrum Mikroelektronik Dresden AG, Ingolstadt, 2013.

[Fr07] R. Frost, D. Rudolph, C. Galke, R. Kothe & H.T. Vierhaus. A Configurable Modular Test Processor and Scan Controller Architecture. In Proc. IEEE International On-Line Testing Symposium, S. 277-284, 2007.

[Ga04] Gätzschmann, U.; Galke, C.; Vierhaus, H. T.: Ein flexibles Verfahren zur Testdaten-Kompaktierung und -Dekompaktierung für den Scan-Test. In Proc. GI/GMM/ITG Workshop Testmethoden und Zuverlässigkeit von Schaltungen und Systemen, 2004.

[Gl12] Gleichner, C.; Vierhaus, H. T.; Engelke, P.: Scan Based Tests via Standard Interfaces. In Proc. IEEE Euromicro Conference on Digital System Design, S. 844-851, 2012.

[IJ13] P1687/D1.62 - IEEE Draft Standard for Access and Control of Instrumentation Embedded within a Semiconductor Device. IEEE Association, 2013.

[JT01] IEEE 1149.1-2001 Standard Test Access Port and Boundary-Scan Architecture. IEEE Association, 2001.

[Ko08] Kothe, R.; Vierhaus, H. T.: A Scan-Controller Concept for Low-Power Scan Tests. Journal of Low-Power Electronics, American Science Publishers, Vol. 4, S. 1-9, 2008.

[Ko10] Kothe, R.; Vierhaus, H. T.: Test Data and Power Reductions for Transition Delay Tests for Massive-Parallel Scan Structures. In Proc. IEEE Euromicro Symposium on Digital Systems Design, S. 283-290, 2010.

[Ra04] Rajski, J.; Tyszer, J.; Kassab, M.; Mukherjee, N.: Embedded Deterministic Test. IEEE Transactions on Computer-Aided Design of Integrated Circuits & Systems, Vol. 23, S. 776-792, 2004.

[Sc06] Schölzel, M.: Automatisierter Entwurf anwendungsspezifischer VLIW-Prozessoren. Dissertation, BTU Cottbus, 2006.

[SH09] Secure Hardware Extension - Functional Specification V1.1, Rev. 439. HIS-Herstellerinitiative Software, 2009.

[VS14] VehicleServicePros.com, D. Kolman. DIANA is Looking into Vehicle Electronic Controls. Juli 2010. http://www.vehicleservicepros.com/blog/10342815/diana-is-looking-into-vehicle-electronic-controls, [Stand 28.04.2015].

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Parity-based Soft Error Detection with Software-based Retry vs. Triplication-based Soft Error Correction - An Analytical Comparison on a Flash-based FPGA Architecture

Gökçe Aydos[1] Görschwin Fey[12]

Abstract: *Field-programmable gate arrays* (FPGAs) are often utilized in space avionics. To protect the FPGA logic against the ionizing radiation effects in space, redundancy in form of concurrent error detection can be used.

In this work, we present a comparative study of a parity-based error detection with software-based retry, and a triple modular redundancy technique on a known flash-based FPGA architecture (Microsemi ProASIC3). We compare critical path delay, circuit area overhead, multiple bit error probability and error correction time penalty. Our analysis shows that a solution based on parity-based error-detection can at least save about half of the resource overhead caused by triplication of the flip-flops if the target circuit can be functionally isolated from the rest of the circuit in the FPGA and if the software supports retransmission of access requests.

Keywords: fault tolerance, FPGA, ProASIC3

1 Introduction

Field-programmable gate arrays (FPGAs) are often utilized in space avionics due to their processing efficiency, reprogrammability, and extensible interface capabilities; providing flexibility for a range of mission requirements. The avionics must be protected from ionizing radiation in space. In the absence of a shield (e.g., magnetic field of the earth), a high energy particle can traverse through a digital circuit and induce significant amount of charge, which can eventually cause unexpected system responses like random signal glitches on the sensor data, but also catastrophic system failures like mission-loss due to a nonresponsive satellite [Pe11a]. Due to the lack of cost-efficient physical access to the space system, the avionics must implement intrinsic fault-tolerance mechanisms based on the mission requirements.

A local corruption of information stored in a node by a single energetic particle is called *single event upset* (SEU). If an SEU is latched by an FF (flip-flop), then it can result in a static bitflip. These errors are not permanent and can be corrected e.g., with a reset, thus they are also called *soft errors* [Pe11b]. Soft errors often happen in the sequential elements of a circuit, due to the latching-window, electrical- and logical-barriers of combinatorics [Li94].

[1] University of Bremen, Reliable Embedded Systems, 28359 Bremen, goekce@cs.uni-bremen.de
[2] German Aerospace Center, Institute of Space Systems, 28359 Bremen, goerschwin.fey@dlr.de

Error detection involves only the discovery of an error, while the error correction takes care of both detecting and recovering the correct information after an upset. This requires a sort of information redundancy in form of *space*, e.g., triplication and voting, or *time*, e.g., processing information three times by a single unit and comparing the results. Often, error correction in an FPGA is implemented with space redundancy, especially as *triple modular redundancy* (TMR). Error correction typically requires more resources compared to error detection, in form of redundancy. In presence of tight constraints, this overhead can turn into a hurdle for fulfilling the design timing closure and area requirements.

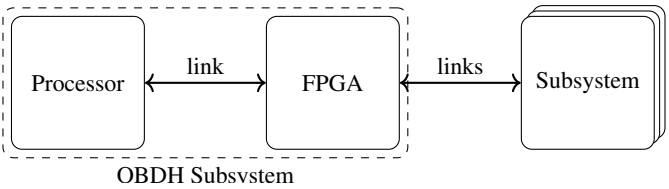


Fig. 1: Overview of an example data handling system. The processor runs the mission software and the FPGA implements interface protocol circuits required by various subsystems on-board of a satellite. The processor uses the FPGA for communicating with the subsystems.

Alternatively, a part of the space redundancy in the FPGA may be eliminated by implementing additional time redundancy, e.g., in software, if the FPGA acts as a co-unit beside an already radiation-hardened processor. An example architecture is depicted in Fig. 1, where the FPGA implements the communication protocol interfaces needed for communicating with the satellite subsystems and the processor runs the mission software. The processor uses the FPGA in a master-slave manner. The FPGA circuit only implements error detection and, in case of an error, the software instructs the FPGA to reprocess the last request. With this collaborative approach, error correction is achieved and the overhead of local error correction is eliminated. This technique will be referred as *error detection with software-based retry* (EDSR). In this paper, a *parity-based* error detection technique is used in the implementation of EDSR.

Parity-based codes and *triplication* are well-known *concurrent error detection* techniques (CED) [NZ98],[Gö08]. Also *error detection with retry* for achieving error correction was proposed, e.g., in [Ni99]. In recent years, one the one hand, partial hardening techniques were proposed due to the relatively high overhead of CED techniques, which selectively harden susceptible parts of the circuit [MT03]. On the other hand, software-based fault-tolerance techniques are also popular due to the flexibility and relatively loose constraints of software, e.g., regarding memory requirements, compared to hardware [Re02],[Go06]. Software- and hardware-based techniques have their tradeoffs, therefore these can also be used together [Re02].

This work applies parity-based error detection with software-based retry and triplication on an example data handling architecture based on a commercially-available flash-based FPGA and provides an analytical comparison of critical path overhead, area overhead, multiple bit error probability, and error correction time penalty. This FPGA is chosen because it is state-of-the-art for space missions (e.g., [Tr14]) and it is optionally available in a special integrated circuit package for space environment. Our contributions are (a) the

discussion of parity-based error detection in the context of the full system stack and (b) the analysis of TMR versus EDSR with respect to space-proven technology.

In the following sections, we firstly explain the TMR and EDSR techniques and the particular implementations which are compared. Then, a reference data processing system is presented, which is used as a testbench for comparing the two implementations. Afterwards, the analytical analysis and its results based on a commercially-available flash-based FPGA architecture are presented. Our analysis shows that the software-based error correction approach can cut down area overhead about 50 % compared to TMR at the expense of extra software runtime, if the target circuit can be functionally isolated from the rest of the FPGA circuit in case of an SEU.

2 Compared Hardening Techniques

In this section TMR and EDSR techniques are described more in detail including their system impacts.

2.1 Triple Modular Redundancy

In TMR, one module is triplicated and the outputs of the three modules are input to a voter, which outputs the majority value. A *module* in this sense can be anything from a whole system to a small functional block or simply a gate. TMR regarding FPGAs can be implemented at various abstraction levels, e.g., at circuit- or gate-level.

There are various TMR techniques based on the reliability requirements of a circuit [Be08]. One of them is the *Local TMR* (LTMR) and is applied on the gate level; a combinational net being registered by an FF is connected to two additional FFs and the outputs of the three FFs are connected to a majority voter. In this work, only SEUs on the FFs are considered. Consequently, the local TMR is used as the compared TMR technique.

TMR detects and corrects a single bit error on an FF locally using a majority voter, hence the TMR techniques can be automatically applied on top of a circuit. This makes TMR functionally transparent to the rest of the system, consequently the circuit mostly does not require a redesign before mapping to an FPGA.

2.2 Error Detection with Software-based Retry

Detection of an error also requires space or time redundancy, but often less redundancy resources than both detection and correction. If the resources on a device are scarce and costly, then implementing a local error correction scheme can become a hurdle. In this case, the error correction can be moved, e.g., to software, if the processing architecture renders it possible. Issuing a non-local error correction requires more recovery time than

a local correction, beginning from the detection until the correction of the error. Nevertheless, if the error rate of the system is low, then a non-local error correction can be practicable.

A well-known error detection technique is *parity-based error detection* (PBED), which adds a parity bit to every data word being stored, e.g., by XORing the data bits and storing the result along with the data word. [NZ98] Upon reading the data word, the parity is calculated again, compared to the stored parity value and in case of a mismatch, an error signal is asserted. Subsequently, an error handler can react and initiate a recovery scheme to correct the error.

After an error, a module must be recovered to an operational state. Often, this is done by resetting the module to its initial state. This in turn leads to a loss of the processing context that must be brought back, which involves periodically backing up the processing context, i.e., checkpointing. If the processing context does not contain any information which is needed for a long time, i.e., when a module regularly falls back to a defined state, then the overhead of checkpointing in the circuit may be eliminated by reissuing a processing request. Examples for such a module are a protocol converter or simply a module which exchanges data between two modules after reformatting data. These modules do not have to store an information for a long time and have a defined state after a chunk of data or a transaction is processed. The example FPGA circuit B presented in Fig. 2 falls also in this category, as it only exchanges data between two modules and moves to its initial state after a request is processed. If an SEU occurs during processing of a request, then the error handler can reset the processing module and flag an error to the processor that a processing request can be reissued, i.e., software-based retry. Alternatively, instead of flagging, the request can be reissued after a nonresponsive timeout. In this case, the time penalty caused by an SEU is negligible, if the FPGA SEU rates during a mission due to space radiation are low.

3 Reference Architecture

We compare the two hardening techniques using a reference model of an on-board data handling unit for a satellite [Tr14]. In the following, important parts of the system are described at the functional level.

3.1 Overview

The on-board data handling unit comprises of two main processing modules: a processor and an FPGA. The processor runs the mission software, which involves communicating with different subsystems on-board of the space system. The communication is done through the FPGA, which acts as an interface component and implements the various communication interfaces needed by the subsystems (e.g., RS232, CAN). Fig. 1 shows an overview of the architecture. We assume that the processor, the communication line between the processor and the FPGA, and the subsystems are sufficiently protected from any

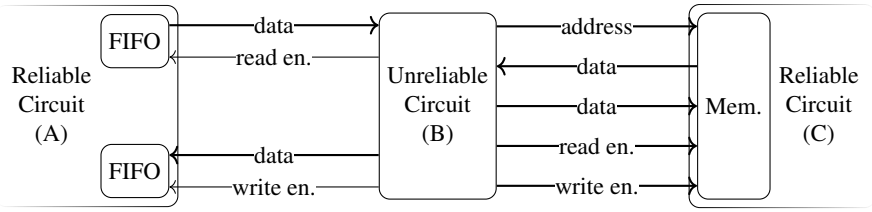


Fig. 2: Simplified model of the FPGA design architecture. The circuit in the middle processes the memory access requests from the left side and responds according to the result of the processed request. The reliable parts are assumed to be immune to SEUs. The unreliable circuit must be hardened on the design level.

single upsets through intrinsic redundancy in ICs and error-correcting code on communication links.

3.2 FPGA Design

From the processor point of view, the FPGA is a remote memory bus, where the implemented link interfaces are memory-mapped. The processor utilizes these interface modules by reading and writing the respective memory areas.

The FPGA model consists of three functional blocks, circuit A, B, and C as shown in Fig. 2. Circuit A serves the memory access requests from the processor to circuit B, which issues memory accesses on circuit C and finally returns the data to the processor using the FIFO interface of circuit A. Circuit C with a memory block inside resembles the memory-mapped interfaces. Reliable circuits A and C in this architecture are assumed to be sufficiently protected against SEUs (e.g., by TMR), whereas the unreliable one must be protected by a soft error hardening technique. The compared hardening techniques will be applied on the unreliable circuit.

The FIFOs and the memory need a single clock cycle for reading or writing a single word, which renders the masking of a single word access operation in a clock cycle (in case of an error) possible.

3.3 Communication Protocol

The communication protocol between the processor and the FPGA consists of two kinds of messages: *request* and *response*. The processor sends memory access requests for a specific address or address interval to the FPGA and the FPGA responds with the according response: In case of a read request, the response carries the data which is requested by the processor. If a write request is issued, the FPGA sends an acknowledge (ACK) response after the write request is complete. A not-acknowledge (NACK) response is sent, if a request cannot be successfully processed. Every request is answered with a response and a second request cannot be sent before the response to the first request has been received.

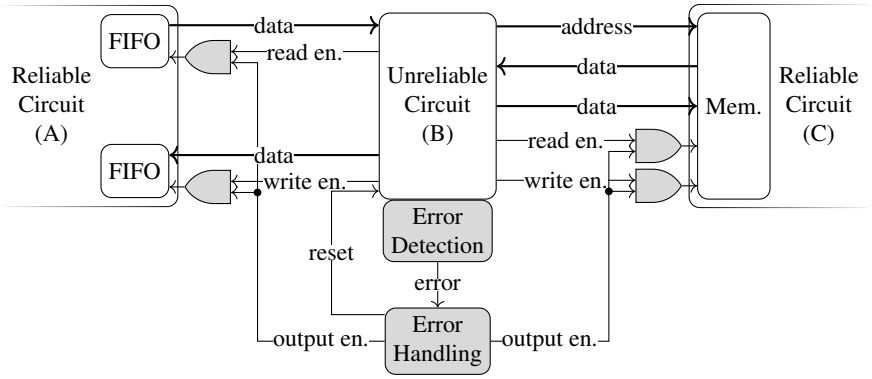


Fig. 3: Parity-based error detection applied on the unreliable circuit. Data redundancy is generated by the parity generation block. The error detection block checks the data integrity. The error handling block generates the recovery and masking signals. The AND gates isolate the unreliable circuit by masking all the control signals which can change the state of the neighbor circuits.

4 Implementation

This section explains how the particular TMR and EDSR using PBED techniques are implemented on the reference architecture.

4.1 Triple Modular Redundancy

In this work, we concentrate on the SEUs in the sequential parts of a circuit, therefore LTMR is implemented as the compared TMR technique. The implementation is straightforward and it does not require additional attention on the hardened circuit, because it can be applied on top of a logical circuit before it is placed and routed for the FPGA.

4.2 Error Detection with Software-based Retry

Fig. 3 shows PBED applied on the reference FPGA design. The *error detection* block continuously generates the data redundancy and checks the integrity of data. If an error is detected, the *error* signal is asserted and the *error handling* block immediately masks the control signals on either side of the unreliable circuit.

The FFs in the unreliable circuit are segmented to groups and for each group one parity FF is introduced. One single group with a parity FF is called a *cluster*. Fig. 4 shows the generic implementation of the error detection in a single cluster. The number of clusters is given by c_{cl} (c: count, cl: cluster). Each cluster contains $s_{\mathrm{cl}} - 1$ user FFs plus one parity FF (s: size). Even parity is generated by XORing the inputs to the user FFs by the $\mathrm{XOR}_{\mathrm{pg}}$. The integrity of the stored bits is checked by the $\mathrm{XOR}_{\mathrm{pc}}$ with s_{cl} inputs and the *cluster error* is generated by each cluster. Finally, c_{cl} cluster error signals are reduced to a single

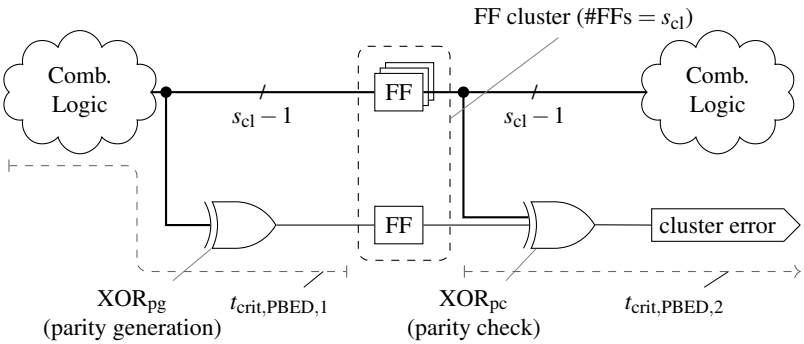


Fig. 4: Implementation of PBED part (a): generation of the cluster error signal.

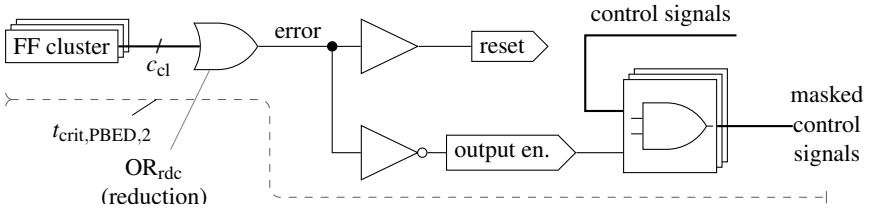


Fig. 5: Implementation of PBED part (b): generation of the global error signal and error handling.

error signal by an OR gate. The reduction of the cluster error signals and subsequent error handling is shown in Fig. 5.

The error handling is done by generating the *reset* and *output enable* signals combinationally using the error signal. The enable signal masks the control signals (i.e., FIFO and memory control signals) of the unreliable circuit. The reset signal recovers the circuit from a possibly erroneous state to its initial state. In the next cycle, the error flag is deasserted and the unreliable module begins data processing again.

If an incomplete or no response is received by the processor in the timeout window, then a recovery procedure is initiated. If an error happens during processing of a read request, then this request is repeated. If an error occurs in the middle of a write transaction, the software cannot know which part of the transaction was completed and the software can synchronize itself by reading these addresses again or simply retry the last transaction. If a write to a memory location triggers an operation (e.g., transmitting a command to a subsystem), then retrying retriggers the last operation, which can be undesirable and dangerous.

In case of such *action-triggering* memory locations, the software can issue single memory write operations only. This has the advantage that every atomic memory write operation is acknowledged separately and the software knows exactly which single memory operation did not succeed. This requirement can be loosened, if a memory area is written which does not trigger an action, i.e., the output of the target system does not change after the

transaction. An example is the transmit buffer of a communication interface module, where the transmit operation must be first triggered by setting a bit in a control register allowing to begin a data transfer to a subsystem. In this case, the processor would first try to write the transmit payload-data to the buffer with a single write request and in the subsequent request the transmission operation would be triggered using an atomic memory access.

5 Analytical Comparison of Needed Resources

In this section, we analytically determine and compare critical path delay, circuit area overhead, multiple bit error probability and discuss the error correction time penalty of the two shown techniques. The circuits are mapped to the Microsemi flash-based radiation-tolerant ProASIC3 FPGA (RT3PE) [Mic13] featuring three input LUTs. This is a known space-proven FPGA and utilized in state-of-the-art on-board-computing systems.

Many of the comparison parameters are dependent on the size of one cluster s_{cl} and the total cluster count in the unreliable circuit c_{cl}. The parameters are determined for $s_{\mathrm{cl}} \stackrel{!}{=} 3^x$ and $c_{\mathrm{cl}} \stackrel{!}{=} 3^y$, where $x, y \in \mathbb{N}$, which fits the RT3PE architecture with three input LUTs.

This selection of input parameters makes the most timing-efficient use of the FPGA area for a specific logic depth. With the increasing count of s_{cl} and c_{cl} more LUTs are needed for parity generation and the reduction of cluster error signals, respectively. With increasing number of LUTs on a critical path, longer delay is introduced on this path. However, the additional delay is only proportional to the logarithm of s_{cl} and c_{cl}. Consequently, the critical path of a benchmark design only changes for different values $x, y \in \mathbb{N}$, leading to such selection of s_{cl} and c_{cl} values. This behavior is visualized in Fig. 6 and explained in Subsection 5.1 more in detail.

In PBED, for each group of $s_{\mathrm{cl}} - 1$ FFs one parity bit is generated. Logic optimization (e.g., logic packing, retiming) and interconnect delays are not considered, which depend significantly on the resource utilization in an FPGA.

In the following, the nominal parameters (i.e., hardening not implemented) are labeled with the subscript $_{\mathrm{nom}}$ and the parameters of the circuits with LTMR and PBED with $_{\mathrm{LTMR}}$ and $_{\mathrm{PBED}}$, respectively. An overhead in a measurement parameter by the applied technique is labeled with the subscript $_{+}$.

5.1 Critical Path Delay

The critical path delay t_{crit} limits the maximum frequency of a design and increases with additional serial logic. In LTMR, every bit must be decoded by a majority voter (MAJ3) before it is propagated to the combinational logic, which causes an extra delay. The subscript $_{\mathrm{pd}}$ stands for propagation delay.

$$t_{\mathrm{crit+,LTMR}} = t_{\mathrm{pd,MAJ3}} \tag{1}$$

In PBED, there are two critical path candidates (see Fig. 4 and 5):

1. The critical path of the nominal circuit plus the parity generation path ($t_{\text{crit,PBED,1}}$)
2. The error detection plus the error handling plus the isolation path ($t_{\text{crit,PBED,2}}$)

The first path delay can be calculated as follows: The parity has to be generated before the combinational signals are registered. The propagation delay of the XOR_{pg} block is called $t_{\text{pd,XOR}_{\text{pg}}}$.

$$t_{\text{crit+,PBED,1}} = t_{\text{pd,XOR}_{\text{pg}}} \tag{2}$$

The error detection path consists of the XOR_{pc}, OR_{rdc}, a NOT gate, and an AND gate (Fig. 4 and 5). The NOT gate and the AND gate can be packed into one LUT, which is called *OR2A*:

$$t_{\text{crit,PBED,2}} = t_{\text{pd,XOR}_{\text{pc}}} + t_{\text{pd,OR}_{\text{rdc}}} + t_{\text{pd,OR2A}} \tag{3}$$

XOR_{pc}, XOR_{pg} and OR_{rdc} are trees of LUTs as shown in Fig. 6. The propagation delay of a *block* with an input size s_{input} is called $t_{\text{pd}}(block, s_{\text{input}})$ and can be calculated by determining the depth d of the tree and multiplying it with the propagation delay of the respective three input macro (e.g., OR3 for an OR block), as the interconnect delays are not considered.

$$\begin{aligned} t_{\text{pd}}(block, s_{\text{input}}) &= d_{block} \cdot t_{\text{pd},macro} \\ &= \lceil \log_3 s_{\text{input}} \rceil \cdot t_{\text{pd},macro} \end{aligned} \tag{4}$$

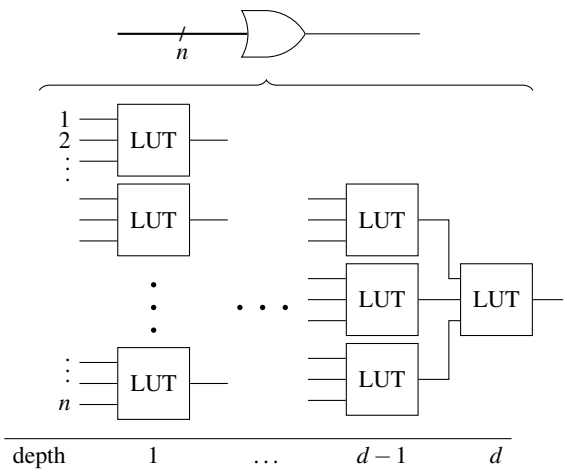


Fig. 6: The figure shows how a gate with n inputs is mapped to an FPGA architecture with three input LUTs. After mapping, a LUT tree with a depth of $d = \lceil \log_3 n \rceil$ is created. Note that if n is not a power of three, then not all the leafs of the tree exist.

With Eq. 4, the propagation delays of the three defined blocks can be calculated:

$$\begin{aligned} t_{\mathrm{pd,XOR_{pg}}} &= \lceil \log_3(s_{\mathrm{cl}} - 1) \rceil \cdot t_{\mathrm{pd,XOR3}} \\ t_{\mathrm{pd,XOR_{pc}}} &= \lceil \log_3 s_{\mathrm{cl}} \rceil \cdot t_{\mathrm{pd,XOR3}} \\ t_{\mathrm{pd,OR_{rdc}}} &= \lceil \log_3 c_{\mathrm{cl}} \rceil \cdot t_{\mathrm{pd,OR3}} \\ &= \left\lceil \log_3 \left\lceil \frac{c_{\mathrm{FF,nom}}}{s_{\mathrm{cl}}} \right\rceil \right\rceil \cdot t_{\mathrm{pd,OR3}} \end{aligned} \tag{5}$$

As $t_{\mathrm{crit,PBED,2}}$ is generated in parallel to the nominal circuit (i.e., not serial like $t_{\mathrm{crit,PBED,1}}$), $t_{\mathrm{crit,PBED,2}}$ stays uncritical up to a certain depth of parity check and reduction blocks. Therefore the parameters s_{cl} and $c_{\mathrm{FF,nom}}$ limit the maximum frequency of the design.

At a junction temperature of 125°Cand worst-case supply voltage 1.14 V, the $t_{\mathrm{pd,MAJ3}}$, $t_{\mathrm{pd,XOR3}}$, $t_{\mathrm{pd,OR3}}$, $t_{\mathrm{pd,OR2A}}$ are 1.14 ns, 1.42 ns, 1 ns and 1 ns[3] respectively [Mic13]. With these data the critical path caused by the FFs and combinational elements can be calculated for various s_{cl} and $c_{\mathrm{FF,nom}}$ parameters.

			$t_{\mathrm{crit+,1}}$ (ns)		$t_{\mathrm{crit,2}}$ (ns)	$Area_+$		$Area_+ : c_{\mathrm{FF,nom}}$	
(x,y)	s_{cl}	$c_{\mathrm{FF,nom}}$	PBED	LTMR	PBED	PBED	LTMR	PBED	LTMR
(1,4)		162			7.84	248	486	153 %	
(1,5)	3	486	1.42	1.14	8.84	734	1458	151 %	300 %
(1,6)		1458			9.84	2192	4374	150 %	
(2,3)		216			8.26	251	648	114 %	
(2,4)	9	648	2.84	1.14	9.26	737	1944	114 %	300 %
(2,5)		1944			10.26	2195	5832	113 %	
(3,2)		234			8.68	260	702	111 %	
(3,3)	27	702	4.26	1.14	9.68	746	2106	106 %	300 %
(3,4)		2106			10.68	2204	6318	105 %	

Tab. 1: Comparison of PBED and LTMR regarding critical path and area overhead.

Table 1 shows the critical path delays $t_{\mathrm{crit+,1}}$ and $t_{\mathrm{crit,2}}$ for various values of the input parameter (x,y). The parameters s_{cl} and $c_{\mathrm{FF,nom}}$ are determined using (x,y), where $s_{\mathrm{cl}} = 3^x$, cluster count $c_{\mathrm{cl}} = 3^y$ and nominal FF count $c_{\mathrm{FF,nom}} = (s_{\mathrm{cl}} - 1) \cdot c_{\mathrm{cl}}$. With increasing depth of $\mathrm{XOR_{pg}}$, $t_{\mathrm{crit+,1}}$ grows for PBED, i.e., every time when s_{cl} reaches a higher power of 3. The additional path delay $t_{\mathrm{crit+,1}}$ of LTMR is independent of the input parameters. For $s_{\mathrm{cl}} = 3$ PBED and LTMR have a similar critical path overhead. PBED has additionally the $t_{\mathrm{crit+,2}}$, which grows with increasing depth of $\mathrm{XOR_{pc}}$ and $\mathrm{OR_{rdc}}$ blocks.

[3] $t_{\mathrm{pd,OR3}}$ and $t_{\mathrm{pd,OR2A}}$ were available neither in the datasheet or macro library documentation. Therefore an estimated value of 1 ns is assumed for these two parameters.

5.2 Circuit Area Overhead

Assuming that the circuit area is proportional to the CLB count, we define the parameter *Area* as the CLB count. For comparison, we are interested in the area overhead $Area_+$, i.e., the CLB cost $c_{\mathrm{CLB}+}$:

$$Area_+ = c_{\mathrm{CLB}+} \tag{6}$$

In ProASIC3 architecture, every *configurable logic block* (CLB)[4] can be either configured as an FF or LUT. Then, the circuit area overhead can be calculated by adding the count of additionally introduced LUTs and FFs:

$$c_{\mathrm{CLB}+} = c_{\mathrm{LUT}+} + c_{\mathrm{FF}+} \tag{7}$$

In the LTMR design, the FFs are triplicated, i.e., two additional FFs are added for each FF:

$$c_{\mathrm{FF}+,\mathrm{LTMR}} = 2 \cdot c_{\mathrm{FF,nom}} \tag{8}$$

LTMR requires one LUT per FF as voter:

$$c_{\mathrm{LUT}+,\mathrm{LTMR}} = c_{\mathrm{FF,nom}} \tag{9}$$

In total, the area overhead for LTMR is:

$$Area_{+,\mathrm{LTMR}} = c_{\mathrm{CLB}+,\mathrm{LTMR}} = 3 \cdot c_{\mathrm{FF,nom}} \tag{10}$$

In PBED, one parity register is needed for a single cluster:

$$c_{\mathrm{FF}+,\mathrm{PBED}} = c_{\mathrm{cl}} \tag{11}$$

In PBED, LUTs are needed for the $\mathrm{XOR_{pg}}$-, $\mathrm{XOR_{pc}}$-, $\mathrm{OR_{rdc}}$-blocks, and OR2A gates for masking the control signals:

$$\begin{aligned} c_{\mathrm{LUT}+,\mathrm{PBED}} =& c_{\mathrm{cl}}\left(c_{\mathrm{LUT,XOR_{pg}}} + c_{\mathrm{LUT,XOR_{pc}}}\right) \\ &+ c_{\mathrm{LUT,OR_{rdc}}} + c_{\mathrm{LUT,OR2A}} \end{aligned} \tag{12}$$

As shown in Fig. 6, a block with n inputs $block_n$ creates a tree, so the needed maximum LUT count for a tree of depth d can be determined by the following formula, assuming that every new level of the tree introduces 3^{depth} LUTs at maximum:

$$\begin{aligned} c_{\mathrm{LUT},block_n,\max} &= \sum_{i=0}^{d_{block_n}-1} 3^i \\ &= \frac{1}{2} \cdot \left(3^{d_{block_n}} - 1\right) \end{aligned} \tag{13}$$

[4] In ProASIC3 terminology, a CLB is called a *tile* or *VersaTile*

Using the formula for depth $d = \lceil \log_3 n \rceil$ (Fig. 6):

$$c_{\text{LUT},block_n,\max} = \frac{1}{2} \cdot (3^{\lceil \log_3 n \rceil} - 1) \tag{14}$$

If n is a power of 3 (e.g., in case of XOR_{pc} and OR_{rdc}), then the equation can be further simplified:

$$\begin{aligned} n \stackrel{!}{=} 3^x, x \in \mathbb{N} &\implies 3^{\lceil \log_3 n \rceil} = n \\ &\implies c_{\text{LUT},block_n} = \frac{1}{2} \cdot (n-1) \end{aligned} \tag{15}$$

If $n+1$ is a power of 3 (e.g., in case of XOR_{pg}), the same amount of LUTs are required. This is due to the fact that a block will in this case contain a single two-input LUT with the rest being three-input LUTs. A two- and a three-input LUT both occupy one CLB, thus the same area.

$$\begin{aligned} n+1 \stackrel{!}{=} 3^x, x \in \mathbb{N} &\implies 3^{\lceil \log_3 n \rceil} = n+1 \\ &\implies c_{\text{LUT},block_n} = \frac{1}{2} \cdot n \end{aligned} \tag{16}$$

With Eq. 15, $c_{\text{LUT,XOR}_{\text{pc}}}$ and $c_{\text{LUT,OR}_{\text{rdc}}}$; and with Eq. 16, $c_{\text{LUT,OR}_{\text{pg}}}$ can be determined. Additionally, there are four OR2As in the PBED implementation. Hence, the Eq. 12 can be rewritten to:

$$\begin{aligned} &c_{\text{LUT}+,\text{PBED}} = \\ &= c_{\text{cl}}\Big(\frac{1}{2} \cdot (s_{\text{cl}} - 1) + \frac{1}{2} \cdot (s_{\text{cl}} - 1)\Big) + \frac{1}{2}(c_{\text{cl}} - 1) + 4 \\ &= c_{\text{cl}}(s_{\text{cl}} - 1) + \frac{1}{2}(c_{\text{cl}} - 1) + 4 \end{aligned} \tag{17}$$

Finally, with Eq. 7, 11 and 17, total area cost for PBED equals to:

$$\begin{aligned} \mathit{Area}_{+,\text{PBED}} &= c_{\text{cl}} + c_{\text{cl}}(s_{\text{cl}} - 1) + \frac{1}{2}(c_{\text{cl}} - 1) + 4 \\ &= c_{\text{cl}}(s_{\text{cl}} + \frac{1}{2}) + 3.5 \end{aligned} \tag{18}$$

$c_{\text{FF,nom}}$ is a main input parameter, therefore it is better to rewrite c_{cl} using $c_{\text{FF,nom}}$:

$$\mathit{Area}_{+,\text{PBED}} = \frac{c_{\text{FF,nom}}}{s_{\text{cl}} - 1}(s_{\text{cl}} + \frac{1}{2}) + 3.5 \tag{19}$$

Table 1 shows the area overhead Area_+ and area overhead caused per FF $\mathit{Area}_+ : c_{\text{FF,nom}}$[5] for various values of s_{cl} and $c_{\text{FF,nom}}$ parameters. PBED area overhead is approximately 59 % of the LTMR area overhead for $s_{\text{cl}} = 3$ and it decreases with increasing s_{cl} and c_{cl}. The LTMR area overhead is independent of the input parameters.

[5] Area overhead Area_+ is related to $c_{\text{FF,nom}}$ instead of the whole design including combinatorics, because the area overhead is only dependent on $c_{\text{FF,nom}}$ and the combinatorics LUT count is arbitrary.

5.3 Multiple bit error probability

We apply the definition of a cluster also on LTMR and define an LTMR cluster as the group of three FFs after triplication, i.e., $s_{\text{cluster,LTMR}} = 3$. LTMR and PBED techniques both are immune against one bit flip in a clock cycle, but not against multiple bit errors, assuming that every FF in LTMR is updated in every clock cycle[6]. In this subsection, we will compare the LTMR and PBED regarding multiple bit error probability, i.e., the probability that an error cannot be detected on the circuit.

If a single particle travels through the circuit, then it can cause single or multiple bit errors dependent on the amount of energy transferred to the circuit and the size of the IC structures. In this analysis, we assume that the CLBs are far enough from each other to consider bitflips as independent events. Therefore, define p as the bit flip probability of a single FF in a clock cycle and assume that p for individual FFs are statistically independent. Then, the multiple bit error (MBE) probability in a single cluster $p_{\text{MBE,cl}}$ can be calculated by:

$$\begin{aligned} p_{\text{MBE,cl}} = & \sum_{i=2}^{s_{\text{cl}}} \binom{s_{\text{cl}}}{i} p^i (1-p)^{s_{\text{cl}}-i} \\ &= 1 - \sum_{i=0}^{1} \binom{s_{\text{cl}}}{i} p^i (1-p)^{s_{\text{cl}}-i} \\ &= 1 - (1-p)^{s_{\text{cl}}} - s_{\text{cl}} \cdot p(1-p)^{s_{\text{cl}}-1} \end{aligned} \tag{20}$$

The last equation assumes that all kinds of multiple bitflips cannot be detected by PBED. In fact, PBED can detect all odd number of bitflips in a cluster, but the probability of multiple bit flips in a cluster greater than 2 is negligible. With $p_{\text{MBE,cl}}$, the multiple error probability of the whole circuit p_{MBE} can be calculated in a similar manner like in the previous equation:

$$\begin{aligned} p_{\text{MBE}} &= \sum_{i=1}^{c_{\text{cl}}} \binom{c_{\text{cl}}}{i} p_{\text{MBE,cl}}^i (1-p_{\text{MBE,cl}})^{c_{\text{cl}}-i} \\ &= 1 - (1-p_{\text{MBE,cl}})^{c_{\text{cl}}} \\ &= 1 - ((1-p)^{s_{\text{cl}}} + s_{\text{cl}} \cdot p(1-p)^{s_{\text{cl}}-1})^{c_{\text{cl}}} \end{aligned} \tag{21}$$

Assuming one year mission in L2 orbit under 1/cm² shielding, a programmed circuit with 5000 FFs on an RTPE3000L FPGA has four SEUs [BSV11]. If this design runs at 20 MHz, then p can be calculated by:

$$\begin{aligned} p &= 4/5000/365/24/60/60/(20 \times 10^6) \\ &\approx 1.27 \times 10^{-18} \end{aligned} \tag{22}$$

Table 2 shows a comparison of multiple bit error probabilities for various s_{cl} $c_{\text{FF,nom}}$ parameters. For $s_{\text{cl}} = 3$, p_{MBE} is approximately the same for PBED and LTMR. Generally,

[6] Otherwise, the bitflips can accumulate and lead to uncorrectable errors.

when $c_{\mathrm{FF,nom}}$ increases, p_{MBE} also increases, but for $s_{\mathrm{cl}} > 3$, PBED is more susceptible to multiple bit errors.

			p_{MBE}	
(x,y)	s_{cl}	$c_{\mathrm{FF,nom}}$	PBED	LTMR
(1,4)	3	162	7.82 E-34	7.82 E-34
(1,5)		486	2.35 E-33	2.35 E-33
(1,6)		1458	7.04 E-33	7.04 E-33
(2,3)	9	216	1.25 E-32	1.04 E-33
(2,4)		648	3.75 E-32	3.13 E-33
(2,5)		1944	1.13 E-31	9.38 E-33
(3,2)	27	234	1.32 E-31	1.13 E-33
(3,3)		702	3.96 E-31	3.39 E-33
(3,4)		2106	1.19 E-30	1.02 E-32

Tab. 2: Comparison of PBED and LTMR regarding multiple bit error probability of one cluster $p_{\mathrm{MBE,cl}}$ and whole circuit p_{MBE}

Overall, $s_{\mathrm{cl}} = 3$ is a reasonable choice for saving significant amount of FPGA resources, and for having as little impact on the critical path as possible. If the maximum frequency is not important, then higher s_{cl} results in less area overhead up to approximately 35 % of the LTMR area overhead.

5.4 Error Correction Time Penalty

In LTMR, the error is corrected in the same clock cycle, but EDSR requires that the error is corrected by the software by repeating the memory access request, which in turn causes additional processing delays. Consequently, the total time penalty is proportional to the error rate during the mission. For example, assuming the same error rate from the Section 5.3 makes the time penalty per year insignificant.

6 Conclusion

The FPGAs used in space applications must be protected against radiation induced errors, which is often done by redundancy. TMR is often used on FPGA designs, which can correct the induced errors locally. If the logic resources are scarce and SEU rate on the FPGA during a mission is low, then the error correction functionality can be shifted to the software, leaving an FPGA circuit only with error detection. We have shown an example architecture which implements PBED with software-based retry and analytically compared the needed resources for the Microsemi ProASIC3 architecture. The results show that significant part of the area overhead caused by the LTMR can be saved by implementing PBED on an FPGA circuit and correcting the errors with time redundancy, i.e., repeating

the instructions to the FPGA in case of an error. The disadvantage of PBED shows up for greater s_{cl} values when the critical path for parity generation grows. Nevertheless, if the maximum frequency is not the first priority in design, then significant area can be saved at a cost of higher p_{MBE}. In certain applications this may be the key to adopt sufficient functionality in a single FPGA component.

Acknowledgment
This work has been supported by the University of Bremen's Graduate School SyDe, funded by the German Excellence Initiative.

References

[Be08] Berg, Melanie: Design for Radiation Effects. Presentation from Military and Aerospace Programmable Logic Devices (MAPLD) Workshop, 2008.

[BSV11] Battezzati, Niccoláo; Sterpone, Luca; Violante, Massimo: Reconfigurable Field Programmable Gate Arrays for Mission-Critical Applications. Springer, chapter 7, 2011.

[Go06] Goloubeva, Olga; Rebaudengo, Maurizio; Reorda, Matteo Sonza; Violante, Massimo: Software-implemented hardware fault tolerance. Springer, 2006.

[Gö08] Gössel, Michael; Ocheretny, Vitaly; Sogomonyan, Egor; Marienfeld, Daniel: New methods of concurrent checking, volume 42 of Frontiers In Electronic Testing. Springer Netherlands, 2008.

[Li94] Liden, P.; Dahlgren, P.; Johansson, R.; Karlsson, J.: On latching probability of particle induced transients in combinational networks. In: 24th International Symposium on Fault-Tolerant Computing (FTCS). pp. 340–349, June 1994.

[Mic13] Microsemi. Radiation-Tolerant ProASIC3 Low Power Spaceflight Flash FPGAs Datasheet, November 2013.

[MT03] Mohanram, K.; Touba, N.A.: Cost-effective approach for reducing soft error failure rate in logic circuits. In: International Test Conference (ITC). volume 1, pp. 893–901, Sept 2003.

[Ni99] Nicolaidis, M.: Time redundancy based soft-error tolerance to rescue nanometer technologies. In: 17th IEEE VLSI Test Symposium. pp. 86–94, 1999.

[NZ98] Nicolaidis, M.; Zorian, Y.: On-Line Testing for VLSI - A Compendium of Approaches. Journal of Electronic Testing Theory and Applications (JETTA), 12:7–20, February 1998.

[Pe11a] Petersen, Edward: Single Event Effects in Aerospace. John Wiley & Sons, chapter 1, 2011.

[Pe11b] Petersen, Edward: Single Event Effects in Aerospace. John Wiley & Sons, chapter 2, 2011.

[Re02] Rebaudengo, M.; Reorda, M.S.; Violante, M.; Nicolescu, B.; Velazco, R.: Coping with SEUs/SETs in microprocessors by means of low-cost solutions: a comparison study. IEEE Transactions on Nuclear Science, 49(3):1491–1495, Jun 2002.

[Tr14] Treudler, Carl Johann; Schröder, Jan-Carsten; Greif, Fabian; Stohlmann, Kai; Aydos, Gökçe; Fey, Görschwin: Scalability of a Base Level Design for an On-Board-Computer for Scientific Missions. In: Proceedings of the Data Systems in Aerospace (DASIA) Conference. 2014.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Detection and Correction of Logic Errors Using Extra Time Slots

Davide Dicorato, Heinrich T. Vierhaus

Abstract: Digital integrated circuits fabricated in nano-technologies have first shown to be more vulnerable to transient errors effects than their predecessors. But they also show effects of stress-induced defects resulting in early life-time failures. In general, power dissipation problems and dielectric stress, due to high field strength, are the main reasons for shortened life-time expectations. On the other hand, system designers require highly reliable and long-time dependable hardware, for example in automotive applications. On-line error detection and-compensation using either codes or, in the more general case, double or triple modular redundancy (DMR and TMR), has been used for decades, but causes higher power dissipation in nano-logic, additional stress, and is therefore no cure in terms of life-time extension. Savings on hardware and power are possible, if resources can be re-allocated to produce local TMR upon demand. However, such techniques may cause sudden signal delays after the detection of errors, which are not easy to handle in synchronous systems. In this paper we present a pseudo-TMR approach, which has little influence on timing in the "good case" and performs a regular error correction within 3 extra clock cycles under error correction without limits on the fault model .

Keywords: Fault tolerant computing, error correction, pseudo-TMR

1 Introduction

New and enhanced fault effects to occur in nano-electronic circuits have been predicted for more than a decade [Ba05,Br04, Fa98,Li09, Lie05]. Intensive research has been done also in the area of aging mechanisms, which harm the switching speed of digital systems and may finally result in delay faults [Al05]. While most research has gone into technologies that may handle either extra delays or transient fault effects [Mi05,Mi06], handling of permanent faults caused e. g. by dielectric breakdowns [Ba09] or by metal migration [Lie05,Lu04] has found much less attention so far. Methods of fast error correction, primarily developed for transient faults such as triple modular redundancy (TMR) or error correcting codes [Pr96,La01], will also work for a limited number of permanent faults, but will not be able to handle transient faults on top of permanent faults in arbitrary combinations. Some research has gone into schemes that provide triple modular redundancy (TMR) by multiple usage of the same elements in time [She09,Ga00] or into schemes that apply TMR selectively only to specific signals [Goe11]. For the repair of permanent faults, repair technologies based on cold redundancy [Ko12,Ko11] are not very fast, since the repair mechanism has to include error detection, fault diagnosis, redundancy allocation and eventually a validation. Such a process may take milliseconds at best and can possibly be part of a system start-up process. Hence on-line available "hot" redundancy is necessary for the detection and compensation of permanent faults that occur in hot operation, followed by an off-line

repair process. In order minimize the total overhead associated with fast error compensation on one side and self repair on the other hand, compatible schemes that can somehow share the overhead have been proposed [Ko12]. The necessary overhead is triplication at least, if all errors have to be compensated in the same clock cycle. If, however, delays of a few extra clock cycles can be accommodated, then a basic scheme of "virtual" triplication is possible, which is based on sharing redundancy between functions that use the same hardware resources [Ko13,Ko14]. Clock management specifically for pipeline structures with on-line detection and correction of delay faults have been dealt with in previous work [Bl04,Bl09, Bl13] and can partly be used for on-line error correction on a wider scale. The basic shortcoming of such methods is the high overhead for the extra circuitry for delay detection and glitch filtering, which causes an overhead of more than 100 % when applied to all signals in a processor design [Bl13].

2 On-Line Error Correction by virtual TMR

A re-configurable logic block (RLB) is considered first. A total of n functions is executed on functional units (FUs) of identical nature. If we protect all functional units by full triplication (fig. 1), the system will be able to handle transient and, with certain limits, permanent faults.

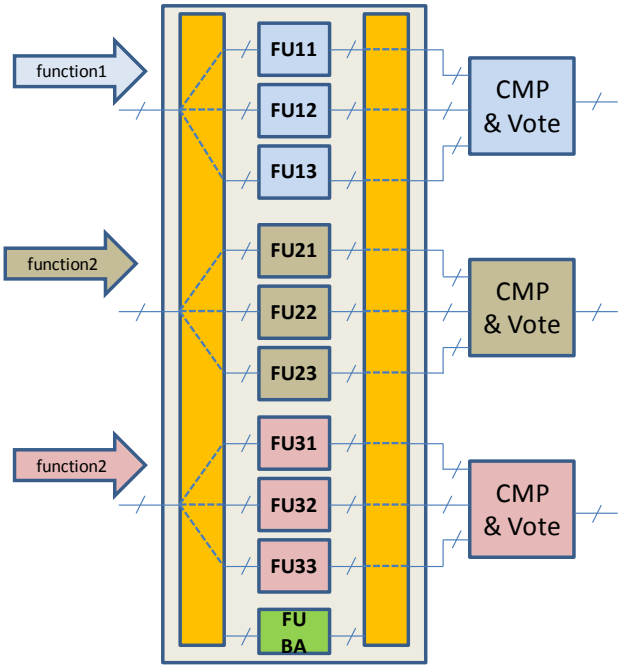


Figure 1. Re-configurable block (RLB) with full TMR and BISR capabilities

With extra power and hardware involved, extra thermal stress may cause wear-out induced faults earlier than in the single system. Then it is still possible to add a limited extra amount of redundancy for the off-line repair of permanent faults. Arrays of switches are needed for the re-location of functions to functional units (FUs). By shifting inputs and outputs between just 2 adjacent FUs, we can use the backup block (FUBA) replace any other defect block (FUkn) in the scheme in an indirect manner. The switching process has to be done off-line in a separate repair mode under either hardware or software control. Hence the existing TMR also has to secure the handling also of permanent faults that occur during normal operation. The drawback of this scheme is the more than triplication in hardware and power. A more cost-effective scheme is shown in figure 2. Now all functions run on duplicate units only. In case of a "fault detect", a third functional unit has to be "borrowed" from a neighboring function, which will then temporarily operate without backup (figure 2).

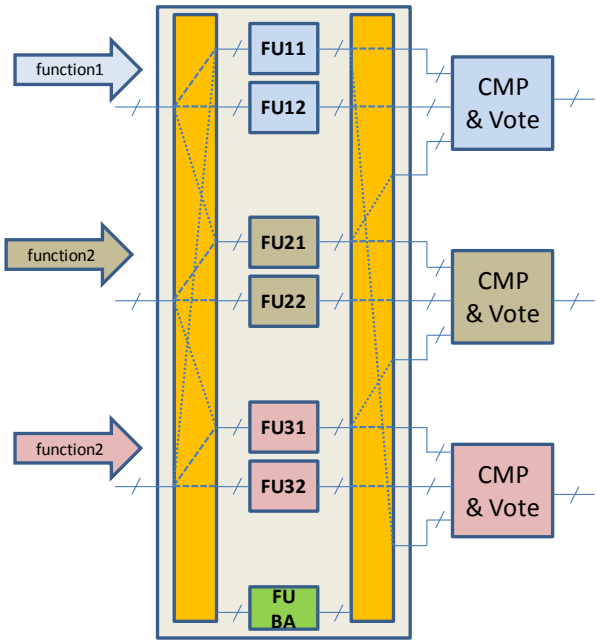


Figure 2. Re-configurable block (RLB) with duplication and selectively "borrowed" TMR

While the overhead now is now essentially determined by duplication, we need an on-line reconfiguration process by switching functions. The comparison of alternatives is shown in Table I. The essential drawback is the demand for extra timing which arises randomly after a "fault detect". Table I compares as alternatives full TMR (fig. 1), a TMR approach with full backup units, where the 3 rd unit is only activated after a "fault detect", and the alternative using borrowed resources from neighboring functions (figure 2).

	Act. TMR	Pass. TMR	Pseudo-TMR
Redund. Overh.	> 200 %	> 200 %	> 100 %
Power Overh.	> 200%	> 100 %	> 100 %
Switching	slow	fast	fast
Timing	regular	irregular	irregular

Table 1. Comparison of TMR-based Architectures

Essentially, it is possible to achieve TMR-capabilities with a reduced overhead in power and hardware, but this goes at the expense of extra clock cycles, which are needed to establish, use, and evaluate temporary TMR. Then the question is, how digital designs can handle "sudden" delays, and how this type of selective TMR can be organized in a robust control scheme. Some previous experimental studies showed that allocation the switching process at inputs and outputs plus the operation of a functional block in one clock cycle is possible, but requires very careful timing and results in reduced clock rates [Mu14]. Therefore we investigated on an alternative approach, where clock cycles need not be stretched in normal operation, but alternatively extra clock cycles are added only in case of fault events.

3 Handling Extra Clock Cycles

Previous research on the detection and correction of delays in pipeline structures has dealt with clocking schemes in several ways [Bl04,Bl09,Bl13]. A global control signal that will "freeze" the clock for one (or more) cycles in order to perform a re-calculation after "fault detect" is one way, which can work with edge-triggered flip-flops and a single clock [Bl04,Bl09].

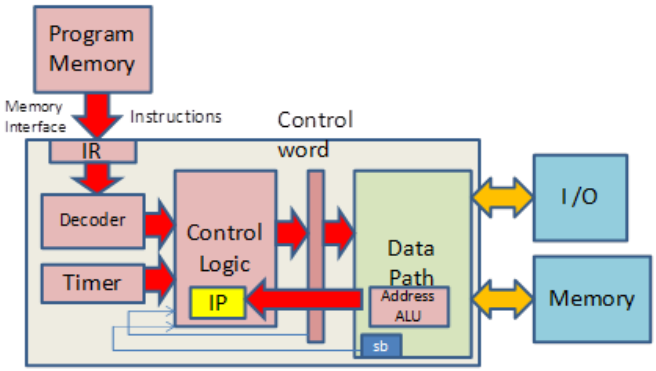


Figure 3. Critical functions for control errors in a typical processor architecture

Specific for the delay fault correction scheme is the problem of “fast” next signal transitions, which may arrive early in the next clock cycle, while the latch that may catch late transitions is still open. The final solution for this problem has been found by the “bubble razor” architecture, which uses a pipeline structure with two non-overlapping clocks, which are alternatively applied to successive pipeline stages [Bl13]. While control logic seems to be less prone to faults induced by path delays [Bl04], it will also need capabilities for on-line error detection and correction. In a typical processor architecture (figure 3), we can define partial functions which are highly control-flow relevant and will therefore need supervision and error correction. These are the instruction memory interface, instruction decoder, the control logic itself, and the ALU needed for address generation. Errors in data path elements not involved in control operations may be tolerated more or less, depending on the application. The methods developed for delay faults in pipelines [Bl04,Bl09,Bl13] are mainly suited for pipelined data paths. Their capability to detect also transient errors has been shown [Bl09], but they are not suited to cover permanent faults. Furthermore, the necessary overhead to protect all signals in a pipeline this way becomes prohibitively high [Bl13]. Assuming delay faults, permanent and intermittent faults as the prime concern, the more universally applicable TMR-based schemes seem to show superior properties. The methods of clock control developed in [Bl04,Bl09,Bl13], however, may also be applied for timing control in partial TMR schemes, which need one or up 3 clock cycles for organization of fault tolerance after an “error detect”. Unlike faults that occur in a pipeline structure, control logic faults should implement a “freeze” function on a local clock network, while a majority vote is performed. The problem then is to organize a repair function which safety executes the majority vote based on “borrowed” resources, while the rest of the system is stalled. The repair process needs to be implemented under hardware control in a robust timing scheme at minimum extra delay for normal operation.

4 Clock and Timing for Borrowed TMR

The simplified version of a pseudo-TMR stage is shown in figure 4. Configuration switches are omitted. We assume master-slave flip-flops at inputs and outputs, composed of D-latches. These latches need specific clock signals upon a “fault detect” condition, which is triggered by comparing the outputs of e. g. blocks FU11 and FU12.Thís signal sets a fault-latch Fl. The outputs from blocks FU11 and FU12 are also latched in the master-latches at the output. The MC-elements, however, pass these signals to the slave latch only, if they are equal, assuming a correct operation. Otherwise the MC-stages show “high impedance”. Then an “error detect” signal from an ongoing calculation must be available, before the output slave latch becomes active and can absorb the faulty input from the master. Subsequently, the master latches of the flip-flops feeding the inputs will have to deliver the same input again for the additional calculations that deliver a correct result by majority vote. This is achieved by manipulating the clock signals applied to the flip-flops at feeding inputs and at the outputs selectively.

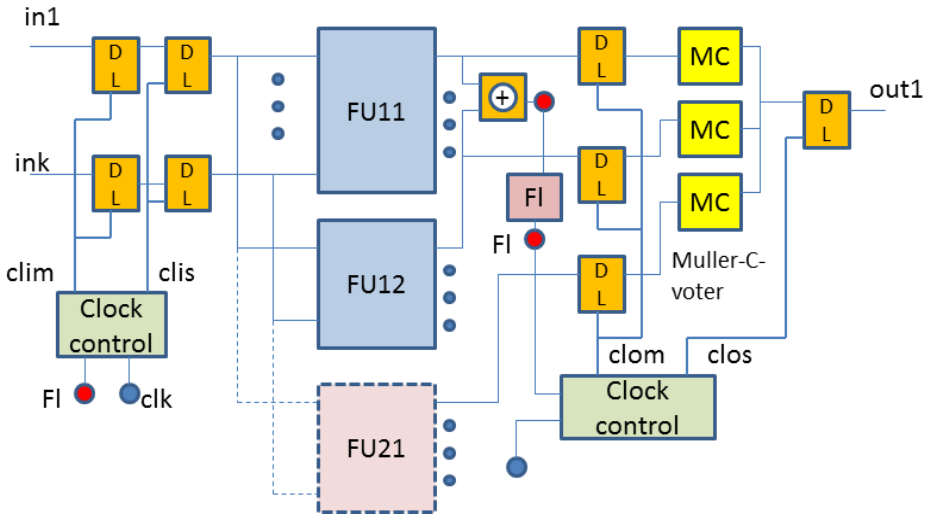


Figure 4. Signal control for fault correction using a third borrowed functional unit (configuration switches omitted)

The "fault detect" signal must arrive early enough to trigger this action. We assume that a fault condition is detected by comparison in clock cycle 0 in time before the high-low-transition. The circuit shown in figure 3 may serve as an example, assuming a faulty behavior in unit FU12.

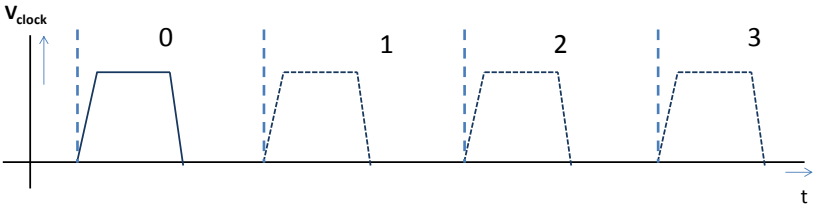


Figure 5. Timing for error detection and compensation

Then the "fault" signal, which is stored in a special fault latch Fl, must prevent false outputs from being transferred to the slaves at the outputs. All slaves receive an inverted "clock high" normally, which is set to low after the fault detect. Next a "fault counter" for detailed control of further timing signals is started, which serves to control the re-configuration and repair action. In phase 1 (figure 5), with all flip-flop outputs frozen, the configuration switches are set in order to re-allocate the third unit from the neighboring function, i. e. FU21 from function 2. This activity should be finished well before the low-high transition of clock phase 1. Then the frozen outputs of the source flip-flops remain attached to the "native" functional units (e. g. FU1, FU2) plus the additional unit FU21, which is the "decider". Then the time from the falling edge of

cycle 1 to the falling edge of cycle 2 is used for re-calculating the triple output and performing a majority vote, e. g. by using a Muller-C-voter. This output is then stored in the output latches right after the high-low transition in phase 2. At the falling edge in clock cycle 2, the correct output is available in the output slave latches. The "fault" signal gets a reset. The next activity, from the falling edge of clock 2 to the rising edge of cycle 3, should be used to re-allocate the switches, e. g. connecting FU21 back to function 2. This can be done while the next input pattern, which was in the inputs of the master-latches at inputs in clock cycle 0, is applied. While the configuration switches are set back, the master latches of the outputs are not active, because the previous pattern that has triggered the fault signal must not be applied again. Only after the falling edge in clock cycle 3, flip-flops at the output should be allowed to take over new data. For controlling the clock signals, we need a single 2-stage counter, which amounts to about 32 transistors. We have also considered to trigger the repair process by "fault" signals only. So far, we use a full comparator only for the outputs of 2 functional units for every function. The detection of a "fault compensated" information would actually mean triple parallel comparison and evaluation by a multi-input NOR gate, triggering "no fault" if the comparison has yielded one in 3 correct outputs.

5 Comprehensive Control for Re-Configuration

So far the re-allocation of resources was done assuming that no functional block in the RLB has permanent faults already. Under realistic conditions, it must be prevented that faulty blocks can be used for the installation of borrowed TMR, since the error correction works only under single fault assumptions within a specific time slot (figure 6).

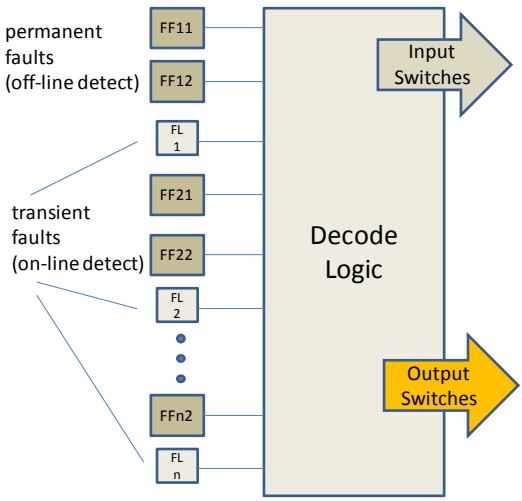


Figure 6. Inclusion of permanent and transient fault states

Hence there must be an additional control of switching inputs and outputs, depending on permanent faults, and units that are needed for permanent repair. The setting of fault bits FF1, FF2… FFn, indicating permanent faults in functional units, can be done during a start-up procedure of the system. If a permanent fault has been detected, one of two parallel functions will execute the associated function permanently. Then a TMR function is no longer possible for this function. Also the support of a neighboring function for "borrowed TMR" is no longer possible. By setting the fault-bits FF1, .. FFn arbitrarily, the system can also be forced into a low-power operating mode. Even then, however, the usage of parallel flip-flops and the MC-element at outputs makes sense, since this circuitry is able to detect and compensate short transient faults in the flip-flops themselves [Mi05]. If a delay line is inserted at the input of two D-latches working in parallel on a single output, also the compensation of short transient faults in the combinational logic blocks is possible [Mi06].

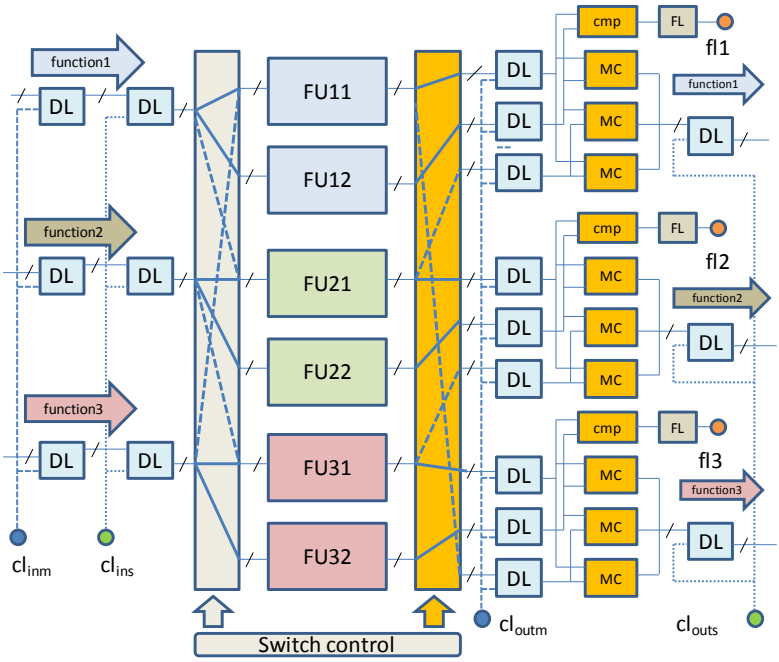


Figure 7. Overall architecture of the re-configurable block

In the overall structure of a re-configurable logic block (fig. 7) we assume that separate clock signals are needed for master- and slave flip-flops at inputs and outputs. At outputs, the slave latches can be composed into a single one, which also serves as the keeper-latch for the MC-elements. We have cl_{inm} for master latches at inputs, cl_{ins} for slave latches at inputs, and, respectively, cl_{outm} and cl_{outs} for the master and slave latches at the outputs. These clock signals can be derived from the normal clock cl and the fault flag bit rather directly.

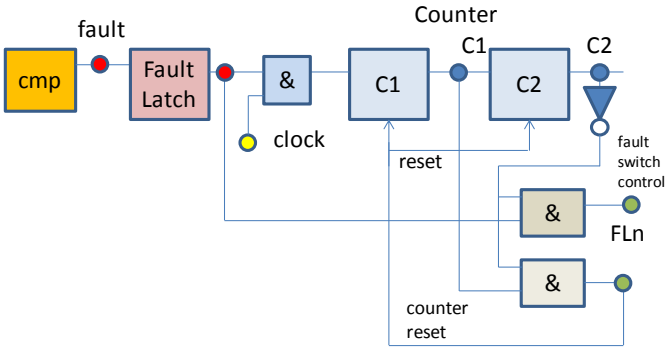


Figure 8. Control signals for fault indication and reset

The FLn signal that controls the setting of input / outputs switches is set to "high" after the arrival of a fault bit from the comparator. It is set back for the passing of the corrected output to the slaves in clock cycle 2. The counter gets a reset in clock cycle 3 (figure 8). At the outputs, the master latches serve to register and hold the outputs of the functional units directly. Results are passed to the output latches (one per bit) only after filtering by the MC-elements, which effectively create a voter, and at the right point of time. Fault effects in the blocks shown in figure 7 will not only influence the affected function, for example function 1 in case of a fault arising in FU12, but possibly also the other functions. Function 2 is in no danger of getting false inputs, while FU 21 is removed temporarily, and the re-attachment of FU21 after error correction is a problem only, if the time for re-setting input switches and for a signal transition through FU21 / FU22 exceeds the clock period. Function 3 is not affected at all. However, looking at global clocking schemes, it may be easier to apply the modified clock with suitable delays to all units. Then a single timing unit (counter and gates) is necessary for the total re-configurable block.

6 Overhead and Limitations

We have published overhead figures for similar architectures in previous work [Ko13,Ko14]. Unlike the architecture in figure 9, which uses separate master and slave latches at outputs, it is also possible to use full master-slave or edge-triggered flip-flops. In any case, the Muller C-elements need to feed a storage device, since their "high impedance" behavior in case of unequal inputs will create a floating node. Other authors have used an extra "keeper latch" at outputs, since then the full flip-flops with master and slave latches are allocated before the MC-elements and benefit from fault correction. Then, however, the keeper latch takes the role as a "single point of failure". The overhead associated with the circuit according to figure 7 is shown in table 2. For comparison, table 3 indicates the overhead for the same type of architecture and basic blocks for full TMR using the same comparators.

Circuit	Trans. Count				Overhead in %					
	Trans. funct.	Ins per f. unit	Outs	Trans. total	b-up	sw	FFs	Cmp/MC	ctrl	total
2-ALU	352	10	6	174	76	13	21	28	16	154
4-ALU	699	14	8	228	82	9	15	20	9	135
8-ALU	1367	24	12	372	86	8	12	17	5	128
16-ALU	2696	38	20	664	88	7	10	15	2	122
32-ALU	5343	70	36	1167	89	6	10	14	1	120

Table 2. Overhead for the RLB-architecture according to figure 9, 3 functions in parallel

Circ.	Transistors		Transistors	Overhead in %		
	unit	total	Voter	back-up	voter	total
2-ALU	352	1056	72	200	20	220
4-ALU	699	2097	96	200	13	213
8-ALU	1367	4101	144	200	10	210
16-ALU	2696	8088	240	200	8.9	208.9
32-ALU	5343	16029	432	200	8.0	208.0

Table3. Overhead for full TMR implementation

Results show that the overhead is dominated by the cost for control and comparison only for relatively small functional units up to about 350 transistors, while in larger entities the cost of duplication is dominating. For comparison, table 4 also shows the cost of a scheme, where at all times all 3 functions are operated on 3 functional units one after the other in a scheme of permutation [Ko13]. This scheme has 3 functional units only and an even larger overhead for flip-flops. Since all calculations are done 3 times regardless fault events, there is a continuous reduction in throughput by a factor of 3, but there is also a constant delay. This scheme has 3 functional units only and an even larger overhead for flip-flops. Since all calculations are done 3 times regardless fault events, there is a continuous reduction in throughput by a factor of 3, but there is also a constant delay.

n = 3			Transistor Count						
BasicBl.	FunctTrans.	In/Outs	Red.Tr.	Sw.Tr.	in FFs	Clog	Decod.	Total	Ovh. %
2-NAND	4/12	2 /1	-	18	180	48	-	258	330
H-Adder	12/36	2 /2	-	18	360	48	-	462	250
F-Adder	30 /90	3 /2	-	27	360	48	-	525	180
2-ALU	352 /1056	10/6	-	90	1080	48	-	2274	69
4-ALU	699 / 2097	14/8	-	126	1740	48	-	4011	61
8-ALU	1367/4101	24 /12	-	216	2760	48	-	7125	57

Table 4. Overhead for time-shared TMR

Unlike more costly schemes that allow for a self repair process based on “cold” redundancy, the architecture shown in figure 11 can handle permanent faults also in multiple units, but after the re-configuration for such purpose, there is no extra resource for handling arbitrary transient faults for a function already infected by a permanent fault. Any probability calculation that we did so far showed that the limitations of reliability and life time are mainly set by elements, which are not protected by duplication. In our case, these elements are the control logic for re-configuration, the switches, and the final slave latches at outputs. Thereby the share of unprotected elements is rather low, compared with more complex repair schemes. Furthermore, there is no extra delay in case of good signals, and the power needed is only a duplication at average, whereby an additional “low power” mode with virtually no extra power, but possible short transient error correction, is also possible.

7 Summary and Conclusions

We proposed a pseudo-TMR-architecture, which is able to detect and correct transient faults and permanent faults equally well. Since apparently delay faults have found most attention recently [Bl04,Bl09,Bl13], we need to compare overhead and limitations. First, delay faults are assumed to occur only on relatively few logic paths, which are specifically identified and protected. Advanced methods [Bl13] will also detect and correct, mostly by re-execution, short transient fault effects in logic and in flip-flops. This comes at relatively high cost per output, since a sophisticated distinction between transient faults and delay faults is implemented. Then the total cost, if applied to all outputs for detection of transient faults and wear-out induced delays, is more than 100 % [Bl13]. The coverage of delay faults in our architecture needs an extra discussion. In case of delay faults, it will not detect delays that occur identically in two functional units such as FU11 and FU12. As, however, delay faults are apparently often corrected by re-execution even under the same timing [Bl13], it is unlikely that delay faults occur as identical double faults. The overhead encompassed here is in the area of about 120 %, but the merit is the universal applicability to different types of faults. Unlike the specific architectures for delay fault correction, this comes at the cost of double power and energy consumption. For the off-line repair of permanent defects by re-configuration, there is a need for some extra overhead for redundancy and switching [Ko14]. Inherently, the scheme holds some promise where several functional units of equal nature can be identified, such as in processor data paths. The application to totally irregular control logic is less economical and would involve a triplication in hardware, but only a duplication in power.

8 Acknowledgment

This work was partly supported by the German Science Foundation (DFG) within the DOZU project.

9 References

[Ba05] R. Baumann: „Soft Errors in Advanced Computer Systems“, IEEE Design and Test of Computers, Vol. 22. No. 3, May / June 2005, pp. 258-266

[Br04] M. A. Breuer, S. K. Gupta, T. M. Mak: “Defect and Error Tolerance in the Presence of Massive Numbers of Defects”, IEEE Design and Test of Computers, Vol.21, No. 3, May / June 2004, pp. 216-227

[Fa98] P. Fang, J. Tao, J. F. Chen, C. Hu: “Design in hot-carrier reliability for high performance logic applications”, Proc. IEEE Custom Integrated Circuits Conference, May 1998, pp. 525-531

[Li09] Y. Li, Y. M. Kim, E. Mintano, D. S. Gardner, S. Mitra, “Overcoming Early-Life Failure and Aging for Robust Systems”, IEEE Design and Test of Computers, Vol. 26, No. 6, Nov. / Dec. 2009, pp. 28-39

[Lie05] J. Lienig. G. Jerke: „Electromigration-Aware Physical Design of Integrated Circuits”, Proc. 18th Int. Conf on VLSI Design, pp. 77-82, 2005

[Al05] M. A. Alam, S. Mahapatra: “A comprehensive model of PMOS NBTI degradation”, Microelectronics Reliability, Vol. 45, No. 1, pp. 71-81, 2005

[Mi05] S. Mitra, N. Seifert, M. Zhang, Q. Shi, K. S. Kim, „Robust System Design with Built-In Soft Error Resilience“, IEEE Computer, Vol. 38, No.2, Febr. 2005, pp. 43-52

[Mi06] S. Mitra, M. Zhang, S. Waqas, N. Seifert, B. Gill, K. S. Kim, „Combinational Logic Soft Error Correction“, Proc. IEEE Int. Test Conf. 2006

[Ba09] M. Bashir, L. Milor, “Modeling Low-k Dielectric Breakdown to Determine Lifetime Requirements”, IEEE Design and Test Vol. 26, No. 6, Nov. / Dec. 2009, pp. 18-27

[Lu04] Z. Lu et al., “Interconnect Lifetime Prediction under Dynamic Stress for Reliability-Aware Design”, Proc. Int. Conf. On Computer Aided Design (ICCAD04), IEEE CS Press, pp. 327-334, 2004

[Pr96] D. K. Pradhan, "Fault Tolerant Computing", Prentice Hall, 1996

[La01] P. G. Lala, „Self-Checking and Fault Tolerant Digital Design“, Morgan Kaufmann Publishers, San Francisco,2001

[She09] X. She, K. S. McElvain, “Time Multiplexed Triple Modular Redundancy for Single Event Upset Mitigation”, IEEE Trans. Nuclear Science Vol.56, No. 4, August 2009, pp. 2443-2448

[Ga00] W. L. Gallagher, E. E. Swartlander, “Fault-Tolerant Newton-Raphson and Goldschmidt Dividers Using Time Shared TMR”, IEEE Trans. Computers, Vol. 49, No. 6, June 2000

[Goe11] M. Augustin, M. Gössel, R. Kraemer, „Implementation of Selective Fault Tolerance with Conventional Synthesis Tools“, Proc. IEEE DDECS 2011, Cottbus, pp. 213-218

[Ko12] T. Koal, M. Ulbricht, H. T. Vierhaus “Combining On-Line Fault Detection and Logic Self Repair”, Proc. 15th IEEE Int. Symposium on Design and Diagnostics of Electronic Circuits and Systems (DDECS 2012), Tallinn, April 2012, Editor J. Raik

[Ko11] T. Koal, D. Scheit, M. Schölzel, H. T. Vierhaus, „On the Feasibility of Built-in Self Repair for Logic Circuits”, Proc. DFT 2011, Vancouver

[Ko13] T. Koal, M. Ulbricht, H. T. Vierhaus. „Virtual TMR Schemes Combining Fault Tolerance and Self Repair”, Proc. Euromicro Conference on Digital Systems Design (DSD 2013), Santander, Sept. 2013 ISBN: 978-3-902457-38-7

[Ko14] Tobias Koal, Mario Schölzel, Heinrich T. Vierhaus: „Combining Fault Tolerance and Self Repair at Minimum Cost in Power and Hardware”, Proc. IEEE DDECS 2014, Warsaw

[Mu14] Sebastian Müller, Tobias Koal, Mario Schölzel, Heinrich T. Vierhaus, „Timing for Virtual TMR in Logic Circuits“, Proc. IEEE On-Line Test Symp. 2014

[Bl04] D. Ernst, D. Blaauw, T. Austin et al., „Razor; Circut-Level Correction of Timing Errors for Low-Power Operation“, IEEE Micro, Nov. / Dec. 2004, pp. 10-20

[Bl09] S. Das, D. Blaauw et al., „Razor II: In Situ Error Detection and Correction for PVT and SER Tolerance, IEEE Journal of Solid State Circuits, Vol. 44., No. 1, January 2009, pp. 32-48

[Bl13] M. Fojtik, D. Fick, Y. Kim, N. Pcikney, D. M. Harris, D. Blaauw, D. Sylvester, “Bubble Razor: Eliminating Timing Margins in an ARM Cortex-M3 Processor in 45 nm CMOS Using Architecturally Independent Error Detection and Correction”, IEEE Journal of Solid State Circuits, Vol. 48, No.1. January 2013

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Möglichkeiten der Modellierung von Fehlern in MLC-Flash-Speichern durch Fehlergraphen

Günther Nieß[1] Thomas Kern[2] Michael Gössel[1]

Abstract: In dieser Arbeit werden beliebige kombinatorische Fehler mit Hilfe eines Fehlergraphen modelliert. Die Knoten des Fehlergraphen sind n-Tupel mit q-ären Komponenten. Zwei Knoten v und v' sind durch eine Kante von v nach v' verbunden, wenn das Wort v durch einen Fehler in das Wort v' verfälscht werden kann. Sollten die Fehler mit bekannten Wahrscheinlichkeiten oder Häufigkeiten auftreten, kann dies durch gewichtete Kanten der Fehlergraphen modelliert werden. Es wird gezeigt, wie ein optimaler fehlererkennender Code mit $m \geq 1$ Prüfbits und $k = n - m$ Datenbits mit einem Fehlergraphen bestimmt werden kann. Für kleine n und m können optimale Lösungen berechnet werden, für größere Blocklängen wird eine effiziente Heuristik vorgeschlagen. Als Beispiel wurden Retention- und Program-Interference-Fehler in Multi-Level NAND-Flash-Speicherzellen modelliert, für die mit Hilfe der vorgestellten Heuristik fehlererkennende Codes bestimmt wurden. Die neuen Codes wurden mit fehlererkennenden Codes für unidirektionale Fehler mit einer beschränkten Auslenkung und mit linearen Blockcodes mit Hilfe experimenteller Ergebnisse verglichen.

Keywords: Fehlermodellierung, Fehlermodell, Fehlererkennung, Flash, MLC, Code

1 Einführung

In dieser Arbeit wird dargestellt, wie $m \geq 1$ optimale Prüfbits eines fehlererkennenden Codes für ein beliebig vorgegebenes Fehlermodell bestimmt werden können. Fehler werden durch einen Fehlergraphen modelliert, in dem die Knoten aus q-ären Worten bestehen. Zwei Knoten v_1 und v_2 sind in dem Fehlergraphen durch eine Kante von v_1 nach v_2 verbunden, wenn ein Fehler des Fehlermodells das Wort v_1 in das Wort v_2 verfälscht. Dieses auf Graphen basierte Fehlermodell erlaubt es, alle kombinatorischen Fehler, die durch unterschiedliche Fehlerursachen ausgelöst werden können, auszudrücken. Fehlerauftrittswahrscheinlichkeiten können den entsprechenden Kanten als Kantengewichte zugeordnet werden.

Für ein gegebenes Fehlermodell werden $m \geq 1$ Prüfbits eines fehlererkennenden Codes mit Hilfe eines Graphen-basierten Algorithmus bestimmt, der durch Entfernen von Knoten die Kantengewichte bzw. die Anzahl der verbleibenden Kanten minimiert. Da die Kanten unterschiedliche Fehler des Fehlermodells entsprechen, ist die Fehlerauftrittswahrscheinlichkeit der verbleibenden nicht erkennbaren Fehler nach dem Bestimmen der optimalen Prüfbits minimal. Es wird weiterhin beschrieben, wie für einen bereits gegebenen fehlererkennenden Codes zusätzliche Prüfbits bestimmt werden können.

[1] Universität Potsdam, AG Fehlertolerantes Rechnen, August-Bebel-Str. 89, 14482 Potsdam, {niess,mgoessel}@cs.uni-potsdam.de
[2] Infineon Technologies AG, Am Campeon 1-12, 85579 Neubiberg, thomas.kern@infineon.com

Die Arbeit ist wie folgt gegliedert. In Abschnitt 2 werden Fehlergraphen eingeführt und es wird dargestellt wie verschiedene Arten von Fehlern ausgedrückt werden können. Abschnitt 3 stellt dar wie Fehlergraphen genutzt werden um fehlererkennende Codes oder m zusätzliche Prüfbits für einen bereits gegebenen Code zu bestimmen. Um auch größere Blocklängen zu bewältigen wird eine Heuristik eingeführt, die eine m-stellige Boolesche Prüfbitfunktion bestimmt. Als mögliche Anwendung der vorgestellten Methode werden in Abschnitt 4 für multi-level Flash-Speicherzellen und Retention-Fehler neu bestimmte fehlererkennende Codes mit bereits bekannten Codes verglichen. Abschnitt 5 schließt die Arbeit mit einer Zusammenfassung ab.

2 Fehlergraph

In diesem Abschnitt wird beschrieben, wie Fehler in Fehlergraphen modelliert werden können. Die Knoten des Fehlergraphen sind n-äre Wörter. Wenn es einen Fehler gibt, der das Wort v in das Wort v' verfälscht, dann wird der Knoten v mit einer gerichteten Kante von v nach v' verbunden. Tritt ein Fehler mit einer Wahrscheinlichkeit p auf, so hat die Kante von v nach v' das Gewicht p. Unter der Annahme, dass es für alle Fehler e, die v in v' verändern auch Fehler gibt, die v' in v verfälschen, ist der Fehlergraph ungerichtet.

Viele Codes wurden zur Korrektur und Erkennung von asymmetrische und unidirektionalen Fehler mit einer beschränkten Auslenkung veröffentlicht [Ca10], [KBE11], [MK12], [EBE13]. Die häufigste Anwendung für diese Codes ist die Fehlerkorrektur und Fehlererkennung in multi-level Flash-Speicherzellen [KBE11], [MK12]. Derartige Speicherzellen haben mehrere Schwellenwerte für Spannungen und somit verschieden programmierbare Speicherzustände. Ein unidirektionaler Fehler mit einer beschränkten Auslenkung verändert den programmierten Wert einer Speicherzelle um einen bestimmten absoluten Betrag in eine Richtung, zunehmend oder abnehmend. Bei asymmetrischen Fehlern mit einer beschränkten Auslenkung werden hingegen nur Fehler betrachtet, bei denen im Voraus schon bekannt ist, ob die Fehler den Wert der Speicherzellen bis zu einem absoluten Betrag erhöhen oder absenken.

In Abb. 1 wird ein Fehlergraph für 2 multi-level Speicherzellen und asymmetrischen Fehlern der maximalen Auslenkung $l = 1$ gezeigt. Dabei werden in den Knoten die möglichen Speicherzustände abgebildet. Jede Speicherzelle kann mit 2 Bits programmiert werden und dadurch die Werte 0 (00), 1 (01), 2 (10) und 3 (11) abspeichern. Der Speicherzustand wird durch die Anzahl der Elektronen in der Speicherschicht definiert. Es gibt hauptsächlich zwei Methoden, die benutzt werden um die programmierte Ladung in den multi-level Speicherzellen in binäre Werte abzubilden. Sie sind in Tab. 1 dargestellt. Die erste Methode

Tab. 1: Multi-level Flash-Speicherzellen und deren Speicherzustände

Gespeicherte Elektronen	kleinste Anzahl			größte Anzahl
Binäres Encoding	11	10	01	00
Gray Encoding	10	11	01	00

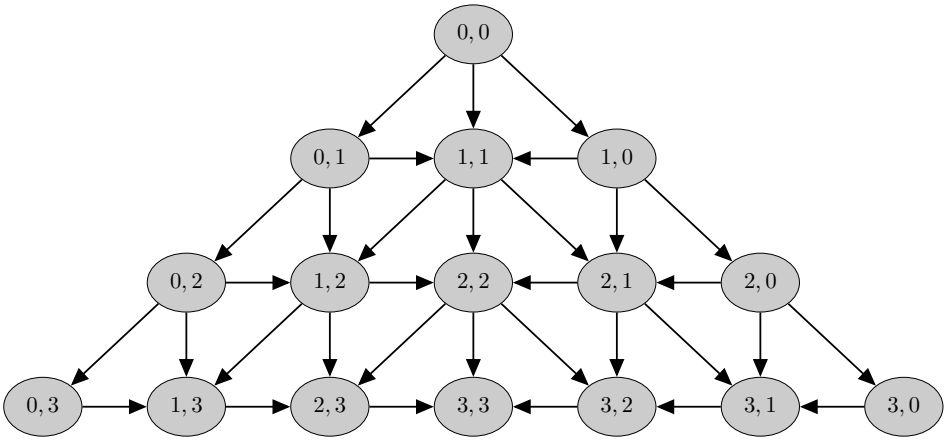


Abb. 1: Fehlergraph für zwei multi-level Speicherzellen mit asymmetrischen Fehlern mit einer beschränkten Auslenkung $l = 1$

benutzt den Gray-Code. Beim Gray-Code unterscheiden sich zwei aufeinander folgende Speicherzustände einer Speicherzelle in nur einem Bit. Wenn man diese Methode benutzt, stellt man sicher, dass eine kleine Veränderung in der Ladungsmenge einer Speicherzelle sich nur in einer kleinen Anzahl von Bit-Fehlern im abgespeicherten Wert auswirkt. Die zweite Methode, das sogenannte binäre Encoding aus Tab. 1, zählt die Schwellwerte einer multi-level Speicherzelle beginnend mit der höchsten Anzahl an gespeicherten Elektronen als 0. Verliert eine Speicherzelle über einen längeren Zeitraum eine geringe Anzahl an Elektronen so verändert sich bei dieser Methode der abgespeicherte Wert ebenso nur um einen kleinen Betrag. Sind zum Beispiel zwei multi-level Speicherzellen mit dem Wert $(0,0)$ programmiert, so halten sie die maximale Anzahl an Elektronen in den Speicherzellen. Tritt ein Fehler mit einer beschränkten Auslenkung $l = 1$ auf, so können die Speicherzellen in die Werte $(0,1)$, $(1,0)$ oder $(1,1)$ verfälscht werden. Dabei verliert im ersten Fall, bei dem der Speicherzustand $(0,0)$ in den Zustand $(0,1)$ verfälscht wird, die zweite Speicherzelle gespeicherte Elektronen, und der Fehler wird im Fehlergraphen aus Abb. 1 durch die Kante abgebildet, die den Knoten $(0,0)$ mit $(0,1)$ verbindet. Werden nun alle asymmetrischen Fehler mit einer Auslenkung $l = 1$ betrachtet, wenn die Speicherzellen mit den Werten $(0,0)$ programmiert wurden, so sind aus dem Fehlergraphen aus Abb. 1 und den drei Kanten, die vom Knoten $(0,0)$ ausgehen, die drei möglichen Fehlerübergänge ersichtlich.

Werden anstatt asymmetrischen Fehlern nun unidirektionale Fehler betrachtet, so werden aus den gerichteten Kanten aus Abb. 1 ungerichtete Kanten. Da für jeden Fehler, der einen Speicherzustand um die Auslenkung $l = 1$ erhöht, wir nun auch den Fehler betrachten, der den Speicherzustand um die Auslenkung $l = 1$ absenkt. Erhöhen wir die Auslenkung auf $l = 2$, kommen zu den Kanten aus Abb. 1 noch zusätzlich die Kanten aus Abb. 2 hinzu. Das jetzt betrachtete Fehlermodell beschreibt ein fehlerhaftes auf- oder entladen von Flash-Speicherzellen um bis zu zwei Stufen. Die schlägt sich in dem grafisch dargestellten Fehlergraphen so nieder, dass nicht nur bestimmte Nachbarknoten durch Fehlerkanten verbunden sind, wie in Abb. 1, sondern dass die Kanten über einen Knoten hinweg zwei

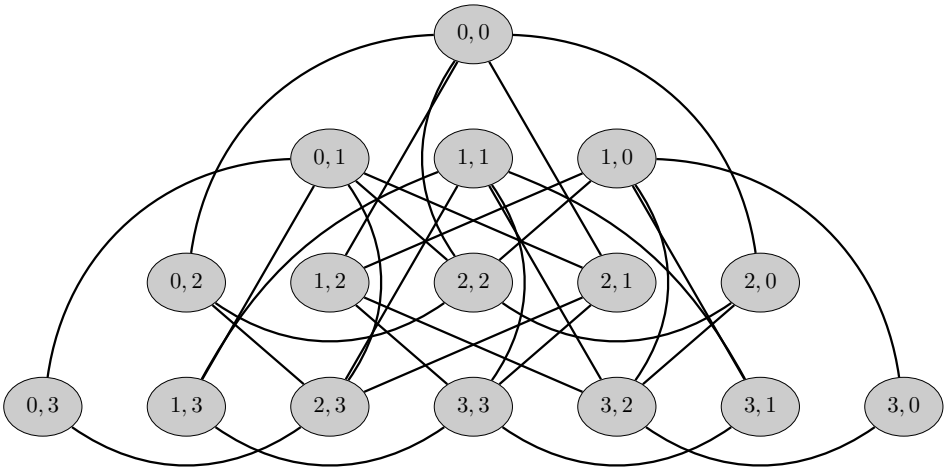


Abb. 2: Fehlergraph mit zusätzlichen Kanten für unidirektionale Fehler mit einer beschränkten Auslenkung $l = 2$

Speicherzustände miteinander verbinden, wie in Abb. 2. Dabei modellieren Kanten, die von oben nach unten, bzw. von außen nach innen verlaufen einen Ladungsverlust in den Speicherzellen und Kanten die von unten nach oben, bzw. von innen nach außen verlaufen eine Ladungszunahme.

Eine ausführliche Analyse von einem multi-level NAND-Flash-Speicherchip machte deutlich, dass Retention-Fehler von den programmierten Speicherzuständen der Flash-Zellen abhängen [Ca12]. Retention-Fehler sind Fehler, bei denen sich der abgespeicherte Wert einer Flash-Zelle über die Zeit verändert. Normalerweise geschieht dies, wenn die programmierte Ladung einer Speicherzelle mit der Zeit über Leckströme entweicht. Eine andere Fehlerart sind die Program-Interference-Fehler. Sie sind nach den Retention-Fehlern die am häufigsten vorkommende Fehlerart [Ca12]. Program-Interference-Fehler treten auf, wenn der Speicherzustand einer Zelle sich verändert, während eine Nachbarzelle beschrieben wird. Dies geschieht durch kapazitive Kopplungseffekte der Nachbarzellen beim Programmieren auf die Speicherzelle, die den Schwellwert verändern. Durch die kleiner werdenden Strukturgrößen und geringeren Abständen zwischen den Zellen wird dieser Effekt zunehmend dominant.

Für die in [Ca12] betrachteten Chips beträgt die Wahrscheinlichkeit von Retention-Fehlern mit ausgeschalteter Fehlerkorrektureinheit (ECC) nach 3 Jahren Betrieb und nach 3.000 Programmier- und Löschzyklen ungefähr $q \approx 10^{-4}$ und die häufigsten Retention-Fehler sind $00 \to 01$, $01 \to 10$, $01 \to 11$ und $10 \to 11$ mit einer relativen Verteilung aller Retention-Fehler von 46%, 44%, 5% und 2%. Die Program-Interference-Fehler besitzen nach demselben Betrieb die Auftrittswahrscheinlichkeit von ungefähr $q \approx 3,5 \cdot 10^{-7}$ und die häufigsten Program-Interference-Fehler sind $11 \to 10$, $10 \to 01$, $10 \to 00$, $11 \to 01$ und $01 \to 00$ mit einer relativen Verteilung aller Program-Interference-Fehler von 70%, 24%, 2,2%, 1,5% und 0,4%. Wenn wir annehmen, dass die Fehlerrate von Flash-Zellen statistisch unabhängig ist und die programmierten Speicherzustände gleichverteilt sind, können wir die

Wahrscheinlichkeiten der Fehlerübergänge von Speicherzellen berechnen. Die verbleibenden drei Prozent der Retention-Fehler, bzw. zwei Prozent der Program-Interference-Fehler wurden nicht genauer in der Arbeit [Ca12] spezifiziert und deshalb hier nicht weiter modelliert.

Abb. 3 zeigt einen Fehlergraphen für 2 Flash-Speicherzellen der mit modellierten Retention- und Program-Interference-Fehlern aus [Ca12] bestimmt wurde. Die Fehlerverteilung wird durch gewichtete Kanten modelliert, die aufgrund der Übersichtlichkeit durch die Linienbreite der Kanten dargestellt wurde, wobei sie der Fehlerauftrittswahrscheinlichkeit des dargestellten Fehlers entspricht. Der Graph stellt ausschließlich Retention- oder Program-Interference-Fehler dar, eine Kombination der Fehler, also dass beispielsweise in der ersten Speicherzelle ein Retention-Fehler und in der zweiten Speicherzelle ein Program-Interference-Fehler aufgetreten ist wird in der Abb. 3 nicht dargestellt. Somit sind alle dargestellten Fehler unidirektional und der Graph ist ein Teilgraph des Fehlergraphen für unidirektionale Fehler mit einer maximalen Auslenkung von $l = 2$. Ein Code der alle unidirektionalen Fehler mit einer maximalen Auslenkung von $l = 2$ erkennt oder korrigiert, kann somit auch alle in Abb. 3 modellierten Fehler erkennen oder korrigieren. Vergleichen wir die Gewichte der Kanten, also die Fehlerauftrittswahrscheinlichkeiten, kann man erkennen, dass einige Kanten mit einer Auslenkung von $l = 2$, z.B. zwischen den Knoten $(01,01)$ und $(01,11)$ ein größeres Gewicht besitzen, als manche Kanten mit einer Auslenkung von $l = 1$, wie z.B. der Fehler zwischen den Knoten $(10,10)$ und $(11,11)$.

Betrachten wir nun Fehler, die beide Speicherzellen betreffen und wobei eine Speicherzelle einen Retention-Fehler und die Andere einen Program-Interference-Fehler aufweist, erhalten wir den Fehlergraph aus Abb. 4. Keine Kante des Graphen aus Abb. 4, die die modellierten Fehler darstellen, ist in dem Fehlergraphen für unidirektionale Fehler mit einer maximalen Auslenkung von $l = 2$ enthalten. Somit werden die dargestellten Fehler von einem Code, der nur unidirektionale Fehler mit einer maximalen Auslenkung von $l = 2$ erkennt oder korrigiert, nicht behandelt und somit nicht erkannt oder eventuell falsch korrigiert. Die modellierten Fehlerauftrittswahrscheinlichkeiten liegen zwischen den Größenordnungen $q \approx 10^{-11}$ und $q \approx 10^{-15}$. Sie unterliegen somit nicht einer so breiten Streuung wie die modellierten Fehler aus Abb. 3. Wie weitere Fehlerarten, wie z.B. Bit-Flips oder Stuck-At-Fehler, durch Fehlergraphen modelliert werden können, wird in [NKG15] beschrieben.

3 Bestimmung von fehlererkennenden Prüfbits

In diesem Abschnitt beschreiben wir, wie zusätzliche Prüfbits für einen fehlererkennenden Blockcode bestimmt werden können. Wir betrachten zwei Fälle.

1. Im ersten Fall werden m optimale Prüfbits c_1, c_2, …, c_m für k Datenbits u_1, u_2, …, u_k bestimmt. Dabei nehmen wir an, dass wir 2^k unterschiedliche k-Tupel codieren können. Wenn die Prüfbits c_1, c_2, …, c_m durch k-äre Boolesche Funktionen f_1, f_2, …, f_m aus den Datenbits u_1, u_2, …, u_k durch $c_1 = f_1(u_1,u_2,\ldots,u_k)$, $c_2 = f_2(u_1,u_2,\ldots,u_k)$, …, $c_m = f_m(u_1,u_2,\ldots,u_k)$ bestimmt sind, bilden die Daten-

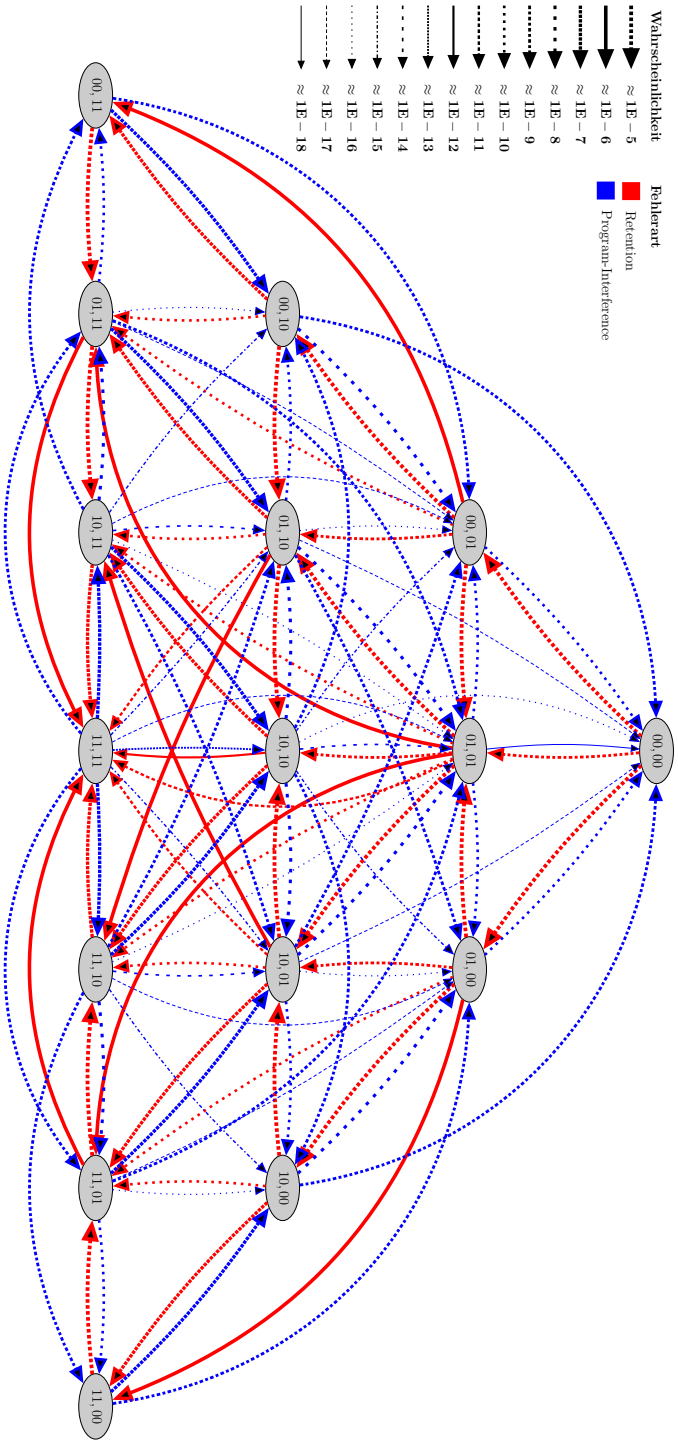


Abb. 3: Fehlergraph für multi-level Flash-Speicher mit unidirektionalen Retention- und Program-Interference-Fehler

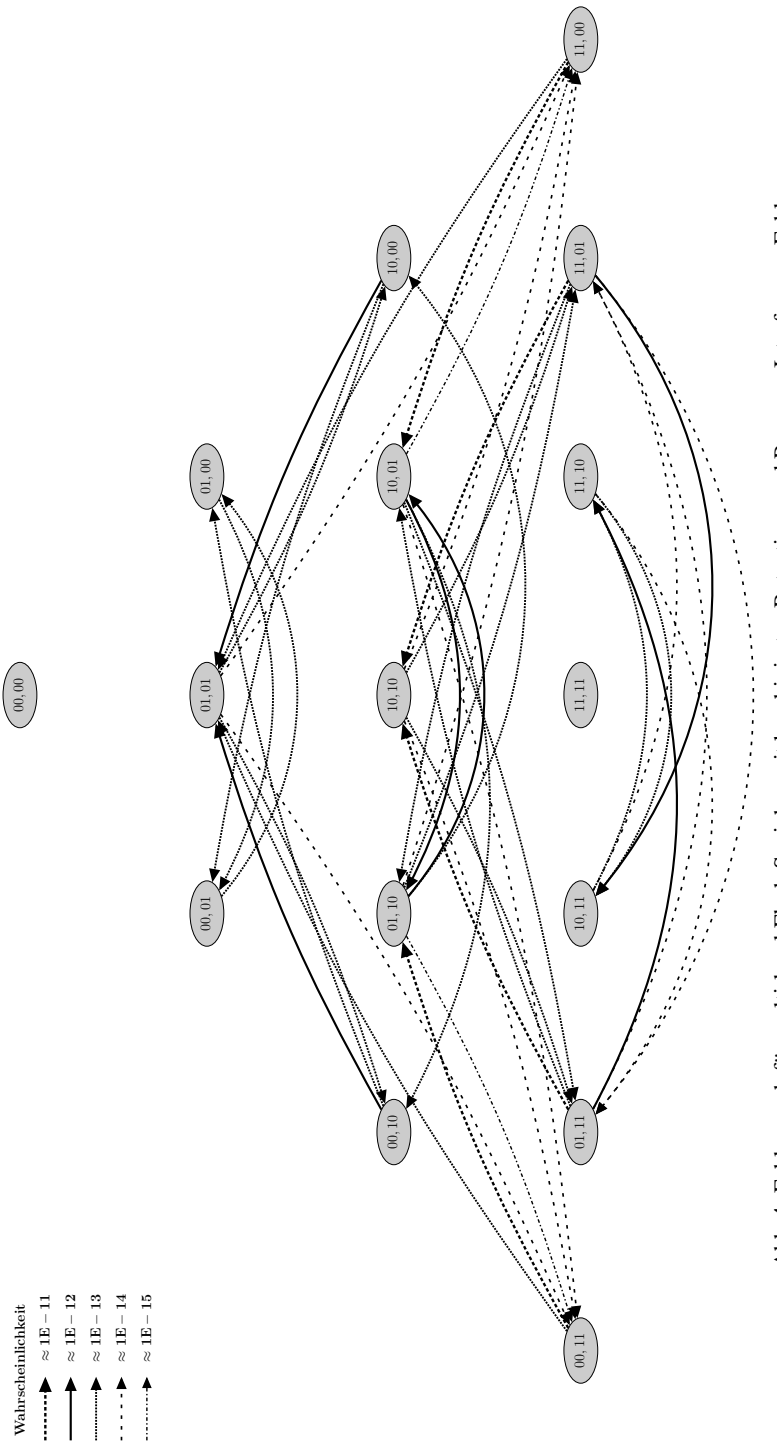


Abb. 4: Fehlergraph für multi-level Flash-Speicher mit kombinierten Retention- und Program-Interference-Fehler

bits mit den entsprechenden Prüfbits $u_1, u_2, \ldots, u_k, c_1, c_2, \ldots, c_m$ ein Codewort des Codes C. Dabei ist der Code C optimal, wenn die Anzahl nicht erkennbarer Fehler oder die Summe der Fehlerauftrittswahrscheinlichkeiten von nicht erkennbaren Fehlern minimal ist.

2. Im zweiten Fall nehmen wir an, dass die Bits $(v_1, v_2, \ldots, v_{n'})$ durch zusätzliche Prüfbits gegenüber Fehlern geschützt werden sollen. Dazu nehmen wir weiterhin an, dass die Bits $(v_1, v_2, \ldots, v_{n'})$ bereits Codewörter eines gegebenen fehlererkennenden Codes C_{out} sind. Durch Hinzufügen von m optimalen zusätzlichen Prüfbits $c_1, c_2, \ldots, c_m$ sollen die Fehlererkennungseigenschaften des Codes C_{out} verbessert werden. Die zusätzlichen Prüfbits werden durch die Funktionen $c_1 = g_1(v_1, v_2, \ldots, v_{n'})$, $c_2 = g_2(v_1, v_2, \ldots, v_{n'})$, $\ldots$, $c_m = g_m(v_1, v_2, \ldots, v_{n'})$ für $(v_1, v_2, \ldots, v_{n'}) \in C_{out}$ bestimmt. Dabei kann der Code C_{out} als äußerer Code betrachtet werden [Fo66], [Bo98]. Sind die Prüfbits $c_1, c_2, \ldots, c_m$ durch Boolesche Funktionen bestimmt, so bilden die Bits $(v_1, v_2, \ldots, v_{n'}, c_1, c_2, \ldots, c_m)$ ein Codewort des neu zu bestimmenden inneren Codes C_{in}.

Betrachten wir zunächst den ersten Fall und nehmen an, dass ein initialer Fehlergraph mit allen 2^n möglichen n-ären binären Knoten gegeben ist, in dem Fehler sowohl in den Daten als auch in den zu bestimmenden Prüfbits modelliert sind.

Für ein gegebenes Datenwort $u = (u_1, u_2, \ldots, u_k)$ bezeichnen wir die Menge aller Knoten mit den gleichen Datenbits u und verschiedenen Prüfbits als Equal-Data-Nodes

$$ED(u_1, u_2, \ldots, u_k) := \{(v_1, v_2, \ldots, v_{k+m}) \in V \mid u_1 = v_1, u_2 = v_2, \ldots, u_k = v_k\}$$

wobei V der Knotenmenge des Fehlergraphen entspricht. Für einen initialen Fehlergraphen besteht die Menge $ED(u)$ aus 2^m unterschiedlichen Wörtern mit 2^m unterschiedlichen Werten für die mögliche Prüfbitbelegung.

Für jeden Knoten $v = (u, c) = (u_1, u_2, \ldots, u_k, c_1, c_2, \ldots, c_m)$ eines Fehlergraphen ist die Menge der komplementären Knoten $CV(v)$ definiert durch

$$CV(u, c) := ED(u) \setminus \{(u, c)\}.$$

Für initiale Fehlergraphen besteht die Menge von komplementären Knoten eines Knotens v aus allen $2^m - 1$ Knoten, die dieselben Datenbits haben wie der Knoten v aber unterschiedliche Prüfbits aufweisen.

Um einen fehlererkennenden Code C zu bestimmen, wird für jedes Datenwort u ein Knoten $v \in ED(u)$ als Codewort $v \in C$ ausgewählt und alle komplementären Knoten $v' \in CV(v)$ werden aus dem Fehlergraphen entfernt. Wenn zum Beispiel der Knoten $(u_1, u_2, \ldots, u_k, 1, 0, \ldots, 0)$ aus dem Fehlergraphen als Codewort ausgewählt wird, erhalten wir die Prüfbitfunktion $1 = f_1(u_1, u_2, \ldots, u_k)$, $0 = f_2(u_1, u_2, \ldots, u_k)$, $\ldots$, $0 = f_m(u_1, u_2, \ldots, u_k)$. Um die Booleschen Funktionen $f_1, f_2, \ldots, f_m$ vollständig zu bestimmen muss für jedes Datenwort u ein Knoten als Codewort ausgewählt werden.

Dies kann auf $(2^m)^{2^k}$ verschiedene Arten erfolgen. Für kleine k und m, $m \cdot 2^k \leq 64$, ist es möglich, alle Fälle zu betrachten und dann den besten oder optimalen Code zu bestimmen.

Für größere k und m wurde eine Heuristik entwickelt. In jedem Schritt des Algorithmus wird mit Hilfe einer Bewertungsfunktion ein Knoten als Codewort ausgewählt und $2^m - 1$ Knoten werden aus dem Fehlergraphen entfernt.

Die Heuristik weist jedem Knoten $v = (u,c) \in V$ ein Gewicht $weight(v)$ zu, welches aus der Summe der Kantengewichte eines Knotens v besteht, die zu dem Knoten v des aktuellen Fehlergraphen führen oder vom Knoten v ausgehen und nicht Knoten der Menge $ED(u)$ enthalten. Ist der Graph nicht gewichtet, so wird jede Kante mit dem Gewicht 1 gezählt. Wird ein Knoten $v = (u,c) \in ED(u)$ als Codewort des Codes C ausgewählt, so verbleiben die Kanten, die mit v und Knoten außerhalb von $ED(u)$ verbunden sind unverändert. Aber alle komplementären Knoten $v' \in CV(u)$ werden zusammen mit ihren Kanten aus dem aktuellen Fehlergraphen entfernt. Um den am besten geeigneten Knoten als Codewort auszuwählen, führen wir den Rang eines Knotens ein

$$rank(v) := \left(\sum_{v' \in CV(v)} weight(v') \right) - weight(v).$$

Die Summe über alle komplementären Knoten $v' \in CV(v)$ beschreibt die Reduzierung der Gewichte, die durch das Entfernen aller komplementären Knoten $CV(v)$ entsteht und der Rang $rank(v)$ berücksichtigt dabei noch die durch die Auswahl verbleibenden Kanten im veränderten Fehlergraphen. Stellen die Kanten die Fehlerauftrittswahrscheinlichkeiten dar, so beschreibt der Rang *rank* eines Knotens v die durch Auswahl des Knotens v zusätzlich gewonnene Fehlererkennungswahrscheinlichkeit minus der Wahrscheinlichkeit das zusätzliche Fehler nicht erkannt werden.

Die Heuristik wählt in jedem Schritt einen Knoten v mit einem maximalen Rang $rank(v)$ als Codewort aus und entfernt seine komplementären Knoten $CV(v)$. In jedem Schritt wird ein neues Codewort des Code C bestimmt, so dass nach 2^k Schritten der Code C vollständig ermittelt ist.

Zum Beispiel sind in Abb. 3 für das Datenwort 01 die Equal-Data-Nodes $ED(01) = \{(01,00), (01,01), (01,10), (01,11)\}$. Die komplementären Knoten des Knoten $(01,01)$ sind $CV(01,01) = \{(01,00), (01,10), (01,11)\}$. Das Gewicht der Retention-Fehler, die den Knoten $(01,01)$ betreffen wird berechnet durch

$$\begin{aligned} weight(01,01) &= 2,12E-9+1,25E-6+1,1E-5+2,2E-10+1,94E-9 \\ &\quad +2,5E-11+2,2E-10+1,15E-5 \\ &= 2,3754525E-5 \end{aligned}$$

und der Rang des Knotens ist

$$\begin{aligned} rank(01,01) &= \left(\sum_{v \in \{(01,00),(01,10),(01,11)\}} weight(v) \right) - weight(01,01) \\ &= 2,375225E-5+2,3752118E-5+2,3750322E-5-2,3754525E-5 \\ &= 4,7500165E-5. \end{aligned}$$

Alle Ränge, die in unserem Beispiel durch die Retention-Fehler des Fehlergraphen aus Abb. 3 betrachten, sind in Tab. 2 aufgeführt.

Tab. 2: Knotenränge der Retention-Fehler des initialen Fehlergraphen aus Abb. 3

Knoten v	$(00,00)$	$(00,01)$	$(00,10)$	$(00,11)$
$rank(v)$	$2,30002E-5$	$2,30000E-5$	$2,30043E-5$	$2,30045E-5$
Knoten v	$(01,00)$	$(01,01)$	$(01,10)$	$(01,11)$
$rank(v)$	$4,75047E-5$	$4,75002E-5$	$4,75050E-5$	$4,75086E-5$
Knoten v	$(10,00)$	$(10,01)$	$(10,10)$	$(10,11)$
$rank(v)$	$2,30043E-5$	$2,30002E-5$	$2,30006E-5$	$2,30038E-5$
Knoten v	$(11,00)$	$(11,01)$	$(11,10)$	$(11,11)$
$rank(v)$	$3,50068E-6$	$3,50004E-6$	$3,50006E-6$	$3,50058E-6$

Entsprechend der Heuristik hat nach Tab. 2 der Knoten $v = (01,11)$ den maximalen Rang und wird demzufolge als Codewort ausgewählt. Seine komplementären Knoten $CV(01,11) = \{(01,00), (01,01), (01,10)\}$ wie auch alle Kanten die mit einem komplementären Knoten $CV(01,11)$ verbunden sind, werden aus dem Fehlergraphen entfernt. Als Ergebnis entsteht

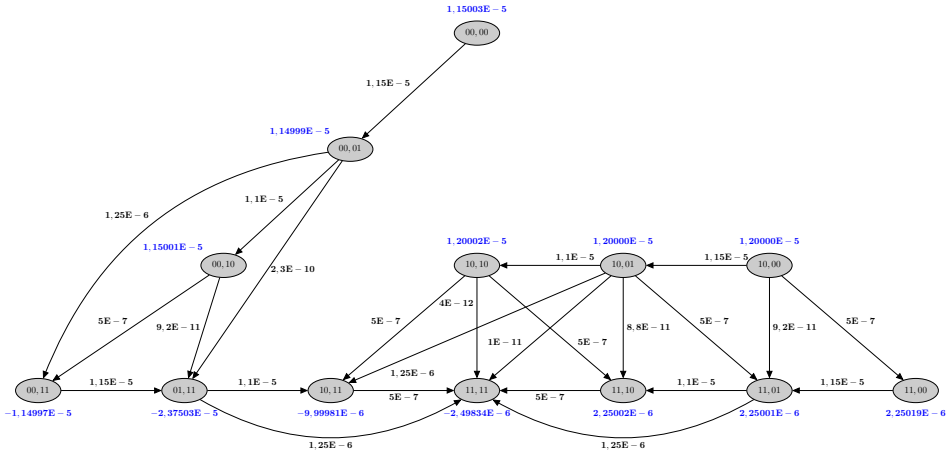


Abb. 5: Gewichteter Fehlergraph nachdem der Knoten $(01,11)$ als Codewort ausgewählt wurde

der in Abb. 5 dargestellte Fehlergraph, bei dem die neu berechneten Ränge über und unter den Knoten abgebildet sind. Nun hat der Knoten $v = (10,10)$ den größten Rang und wird deshalb als Codewort gewählt. Anschließend entfernt der Algorithmus alle Kanten die mit den komplementären Knoten $CV(10,10) = \{(10,00), (10,01), (10,11)\}$ verbunden sind und die komplementären Knoten $CV(10,10)$ selbst aus dem aktuellen Fehlergraphen. In dem resultierenden Fehlergraphen hat der Knoten $(00,00)$ den maximalen Rang $rank(00,00) = 1,15003E-5$ und wird als Codewort gewählt. Nachdem alle komplementären Knoten $CV(00,00) = \{(00,01), (00,10), (00,11)\}$ und die mit den Knoten verbundenen Kanten entfernt wurden, ergibt sich der in Abb. 6 dargestellte Fehlergraph. Der Fehlergraph aus Abb. 6 führt dazu, dass die Knoten $(11,01)$ und $(11,00)$ beide

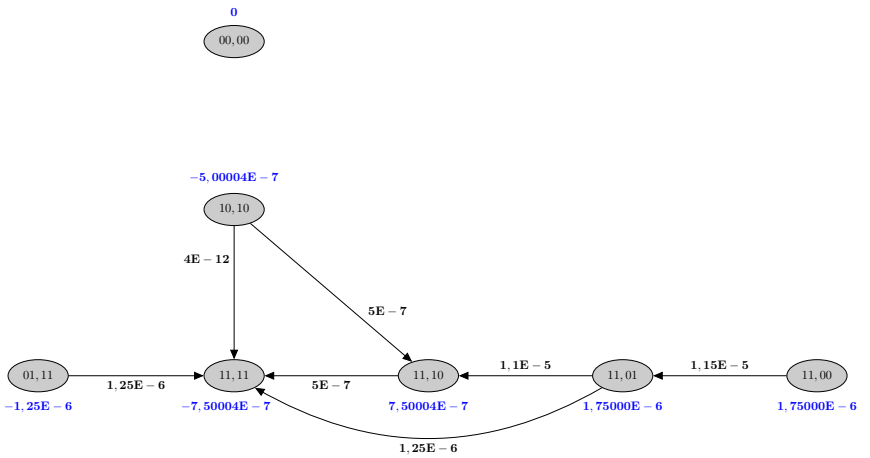


Abb. 6: Gewichteter Fehlergraph nachdem der Knoten $(00,00)$ als Codewort ausgewählt wurde

einen maximalen Rang von $1,75000E-6$ besitzen und wir einen der Knoten als Codewort wählen können. Nachdem ein Knoten als Codewort ausgewählt wurde und seine komplementären Knoten und die mit deren Kanten vom Fehlergraphen entfernt wurden, erhalten wir einen fehlererkennenden Code z.B. $C = \{(00,00), (01,11), (10,10), (11,00)\}$ bei dem die Prüfbitfunktionen f_1 und f_2 vollständig bestimmt sind.

Der zweite Fall ist ähnlich, mit dem Unterschied, dass nur Fehler betrachtet werden, die ein Codewort des inneren Codes C_{in} in ein Codewort des inneren Codes C_{in} verfälschen. Fehler die durch den äußeren Code erkannt werden, können vom initialen Fehlergraphen durch Entfernen der Knoten $(v_1, v_2, \ldots, v_{n'}, c_1, c_2, \ldots, c_m)$ bei denen die ersten n' Bits nicht einem Codewort des äußeren Codes $v \notin C_{out}$ entsprechen nicht weiter betrachtet werden. Die Implementierung des Algorithmus wird in [NKG15] beschrieben.

4 Fehlererkennung für Multi-Level NAND-Flash-Speicher

In diesem Abschnitt werden Codes, die mit der vorgestellten Heuristik erzeugt wurden, mit bereits bekannten fehlererkennenden Codes verglichen. Als beispielhaftes Anwendungsgebiet betrachten wir multi-level NAND-Flash-Speicher mit Retention- und Program-Interference-Fehler die anhand von experimentellen Ergebnissen modelliert wurden [Ca12]. Der Fehlergraph für zwei Speicherzellen setzt sich aus den Graphen aus Abb. 3 und Abb. 4 zusammen. Das Fehlermodell wurde am Ende des Abschnitts 2 beschrieben.

Unidirektionale und asymmetrische fehlererkennende Codes für Speicherzellen, die mehr als $q > 2$ Werte speichern können, werden mit Hilfe des Restklassenrings $\mathbb{Z}/q\mathbb{Z}$ definiert. In praktischen Anwendungen werden hauptsächlich die Werte $q = 2^a$ mit $a \in \{2,3\}$ benutzt. Die Codes besitzen normalerweise ganzzahlige Blocklängen mit $K = k/a$ Datenzellen und $M = m/a$ redundante Prüfbitzellen. Es wird das binäre Encoding von Tab. 1 benutzt um die gespeicherten Elektronen von Flash-Speichern zu interpretieren. Es wird demzufolge ein binäres Codewort $v = (u_1, u_2, \ldots, u_k, c_1, c_2, \ldots, c_m)$ mit k Datenbits

und m Prüfbits mit dem Restklassenring $\mathbb{Z}/2^a\mathbb{Z}$ interpretiert, so dass wir Codewörter der Form $V = (U_1, U_2, \ldots, U_K, C_1, C_2, \ldots, C_M)$ mit $U_1, U_2, \ldots, U_K, C_1, C_2, \ldots, C_M \in \mathbb{Z}/2^a\mathbb{Z}$ betrachten.

Der optimale systematische fehlererkennende Code für unidirektionale Fehler in $\mathbb{Z}/q\mathbb{Z}$ benötigt $\lceil \log_q K(q-1)+1 \rceil$ Prüfsymbole und wird durch

$$c = \sum_{i=1}^{K} (q - 1 - U_i)$$

berechnet, wobei U_1, U_2, …, $U_K \in \mathbb{Z}/q\mathbb{Z}$ den abgespeicherten Werten der Datenzellen entsprechen und die Prüfsymbole $C_1, C_2, \ldots, C_M$ die q-äre Repräsentation von c darstellen [BP82]. In dem beschriebenen Fall von $q = 2^2$ wird für $K = 1$ Datenzelle eine redundante Speicherzelle für die Prüfbits benötigt. Für $K \leq 5$ Datenzellen werden zwei redundante Zellen benötigt um alle unidirektionalen Fehler erkennen zu können. Wenn wir für mehr als $K > 5$ Datenzellen $M = 2$ Prüfsymbole benutzen möchten, empfiehlt sich der Code mit den Prüfsymbolen

$$c = \left(\sum_{i=1}^{K} (q - 1 - U_i) \right) \mod q^2. \tag{1}$$

Der Code erkennt bis zu $t = \frac{q^2}{l} - q$ asymmetrische Fehler mit einer maximalen Auslenkung l [EBE13]. Wir möchten zunächst Codes mit einer redundanten Speicherzelle vergleichen. Diese Codes sollen die am häufigsten vorkommenden Retention- und Program-Interference-Fehler in Flash-Speichern erkennen. Demzufolge können wir zwei Prüfbits oder ein Prüfsymbol im Restklassenring $\mathbb{Z}/4\mathbb{Z}$ definieren. Den Konstruktionsverfahren in [BP82] und [EBE13] folgend, können wir ein Prüfsymbol durch

$$c = \left(\sum_{i=1}^{K} (3 - U_i) \right) \mod 4 \tag{2}$$

bestimmen und es zur Fehlererkennung nutzen.

Eine andere Methode um multi-level Flash-Speicher zu schützen ist, das Gray-Encoding aus Tab. 1 einzusetzen um die Menge der gespeicherten Elektronen in Speicherzellen zu interpretieren und mit einem linearen Blockcode die Fehlererkennung umzusetzen. Der lineare fehlererkennende Code sollte Fehler erkennen, die eine geringe Anzahl von Bit-Fehlern aufweisen. Mit $m = 2$ Prüfbits kann der Code mit

$$\begin{aligned} c_1 &= u_1 \oplus u_2 \oplus u_3 \oplus \ldots \oplus u_k \\ c_2 &= u_1 \oplus u_3 \oplus u_5 \oplus \ldots \oplus u_{k-1} \end{aligned} \tag{3}$$

für die Datenbits u_1, u_2, …, u_k bestimmt werden. Das Prüfbit c_1 entspricht hierbei der Gesamtparität und die XOR-Summe von c_2 enthält ein Bit aus jeder Flash-Zelle. Der Code kann somit alle Fehler erkennen, die nur eine Speicherzelle betreffen oder Fehler die mehrere Speicherzellen betreffen und eine ungerade Anzahl an Bits verfälschen.

Die zwei oben aufgeführten Verfahren werden in Tab. 3 mit den experimentellen Ergebnissen der in Abschnitt 3 vorgeschlagenen Heuristik verglichen. Dabei ist die Tab. 3 in drei

Tab. 3: Die Fehlererkennung von verschiedenen Codes für Multi-Level NAND-Flash-Speicher mit Retention- und Program-Interference-Fehlern

Codewortlänge (n)		4	6	8	10	12	14
# modellierten Fehler		153	2.133	28.305	370.269	4.822.713	62.732.133
Fehlerwahrscheinlichkeit pro Wort		$2{,}0069E-4$	$3{,}0102E-4$	$4{,}0134E-4$	$5{,}0165E-4$	$6{,}0195E-4$	$7{,}0224E-4$
Wahrsch. eines nicht modellierten Fehlers		$6{,}0133E-6$	$9{,}0199E-6$	$1{,}2027E-5$	$1{,}5033E-5$	$1{,}8040E-5$	$2{,}1046E-5$
# Datenzellen (K)		1	2	3	4	5	6
Heuristik	Erkannte Fehler	98,039%	96,906%	96,283%	96,088%	95,696%	95,406%
	Wahrsch. der Erkennung	$1{,}9468E-4$	$2{,}9200E-4$	$3{,}8931E-4$	$4{,}8640E-4$	$5{,}8354E-4$	$6{,}8009E-4$
	Wahrsch. von nichterkannten Fehlern	$1{,}0963E-11$	$2{,}6273E-10$	$3{,}4741E-9$	$2{,}1708E-7$	$3{,}7157E-7$	$1{,}0973E-6$
Unidirektional	Erkannte Fehler	94,771%	94,374%	93,937%	93,800%	93,766%	93,755%
	Wahrsch. der Erkennung	$1{,}9468E-4$	$2{,}9200E-4$	$3{,}8931E-4$	$4{,}8662E-4$	$5{,}8391E-4$	$6{,}8119E-4$
	Wahrsch. von nichterkannten Fehlern	$3{,}0015E-11$	$6{,}4497E-11$	$1{,}2899E-10$	$2{,}1506E-10$	$3{,}2272E-10$	$4{,}5198E-10$
Linear	Erkannte Fehler	95,425%	94,327%	93,913%	93,803%	93,766%	93,755%
	Wahrsch. der Erkennung	$1{,}9468E-4$	$2{,}9200E-4$	$3{,}8931E-4$	$4{,}8661E-4$	$5{,}8390E-4$	$6{,}8117E-4$
	Wahrsch. von nichterkannten Fehlern	$8{,}9907E-11$	$2{,}3687E-9$	$5{,}1636E-9$	$9{,}0318E-9$	$1{,}3973E-8$	$1{,}9987E-8$
# Datenzellen (K)			1	2	3	4	5
Heuristik	Erkannte Fehler		99,953%	99,855%	99,774%	99,766%	99,733%
	Wahrsch. der Erkennung		$2{,}9200E-4$	$3{,}8931E-4$	$4{,}8662E-4$	$5{,}8391E-4$	$6{,}8119E-4$
	Wahrsch. von nichterkannten Fehlern		$2{,}7367E-12$	$4{,}8717E-12$	$4{,}9343E-11$	$4{,}2058E-11$	$4{,}8925E-11$
Unidirektional	Erkannte Fehler		99,625%	99,576%	99,538%	99,529%	99,527%
	Wahrsch. der Erkennung		$2{,}9200E-4$	$3{,}8931E-4$	$4{,}8662E-4$	$5{,}8391E-4$	$6{,}8119E-4$
	Wahrsch. von nichterkannten Fehlern		$7{,}5029E-12$	$1{,}1421E-11$	$2{,}2839E-11$	$3{,}8062E-11$	$5{,}7087E-11$

Abschnitte gegliedert. Der erste Abschnitt beschreibt die modellierten Fehler, im zweiten Abschnitt werden die drei unterschiedlichen Codes mit einer redundanten Speicherzelle verwendet und im unteren Teil der Tabelle werden zwei redundante Speicherzellen zur Fehlererkennung genutzt.

Die erste Zeile führt die Blocklänge n des Codes in Bits auf, entsprechend umfasst ein Codewort $n/2$ Speicherzellen. In der nächsten Zeile kann man die Anzahl der modellierten Fehler finden. Dabei entspricht ein Fehler einer Kante im Fehlergraphen. Ein solcher Fehler ist das Ergebnis, wenn sich eine oder mehrere Speicherzellen unbeabsichtigt auf- oder entlädt. In der dritten Zeile der Tab. 3 wird die Wahrscheinlichkeit aufgeführt, dass ein Retention- oder ein Program-Interference-Fehler innerhalb eines Codeworts aufgetreten ist. Da die Arbeit [Ca12] drei Prozent der Retention- und zwei Prozent der Program-Interference-Fehler nicht weiter spezifiziert, wird in der nächsten Zeile die Wahrscheinlichkeit, dass ein nicht modellierter Fehler auftritt aufgelistet. Die ersten vier Zeilen der Tab. 3 beschreiben also das benutzte Fehlermodell und in unserem Falle den initialen Fehlergraphen.

In den folgenden drei Zeilen (5, 6, 7) werden die experimentellen Ergebnisse der neuen fehlererkennenden Codes beschrieben, die mit Hilfe der vorgestellten Heuristik erzeugt wurden. Die erste Zeile zeigt wie viele der modellierten Fehler mit Hilfe der Heuristik erkannt werden können. In der nächsten Zeile wird die Wahrscheinlichkeit, dass ein Fehler aufgetreten ist und erkannt werden kann aufgeführt. Und in der letzten Zeile der Fehlererkennungsergebnisse der beschriebenen Heuristik wird die Wahrscheinlichkeit dargestellt, dass ein Retention- oder Program-Interference-Fehler auftritt und nicht durch den vorgeschlagenen Code erkannt werden kann. Der fehlererkennende Code für unidirektionale Fehler, der für eine redundante Speicherzelle durch die Gleichung (2) bestimmt wird, wird in Tab. 3 als Unidirektional bezeichnet und seine Ergebnisse werden in den Zeilen 8, 9 und 10 aufgeführt. Die Ergebnisse der linearen Codes sind in den Zeilen 11, 12 und 13 aufgelistet, dabei wird die Menge von Elektronen einer Speicherzelle mit Hilfe des Gray-Codes interpretiert und mit dem binären fehlererkennenden Code aus den Gleichungen (3) werden Fehler erkannt. Im unteren Abschnitt der Tabelle wird die Gleichung (1) aus [EBE13] für die Codes zur Erkennung unidirektionaler Fehler genutzt und mit Codes verglichen, die mit Hilfe der in dieser Arbeit vorgestellten Heuristik erstellt wurden.

Tab. 3 zeigt, dass die auf unidirektionale Fehler spezialisierten Codes in vielen Fällen die kleinste Wahrscheinlichkeit aufweisen, dass ein modellierter Fehler nicht erkannt werden kann. Die mit der vorgestellten Heuristik bestimmten Codes können hingegen die größte Anzahl der modellierten Fehler erkennen, was aus den Prozenten der erkannten Fehler ersichtlich ist. Wenn wir die ganze Tabelle in Augenschein nehmen, können wir erkennen, dass für alle untersuchten Codes die Wahrscheinlichkeit, dass ein nicht modellierter Fehler auftritt um Größenordnungen höher ist, als die Wahrscheinlichkeit, dass ein modellierter Fehler auftritt und nicht erkannt werden kann.

5 Zusammenfassung

In dieser Arbeit wurde gezeigt, wie Fehler durch Fehlergraphen adäquat beschrieben werden können. Als Anwendungsbeispiel wurde das unidirektionale Fehlermodell mit beschränkter Auslenkung einem Fehlermodell gegenübergestellt, dass aus experimentellen Ergebnissen für multi-level Flash-Speicher nach [Ca12] abgeleitet wurde.

Es wurde demonstriert, wie das Bestimmen von Prüfbits eines optimalen fehlererkennenden Codes mit Hilfe eines einfachen graphentheoretischen Algorithmus erfolgen kann, der die Anzahl von Kanten in einem Fehlergraphen minimiert. Zur Berechnung der Prüfbits für größere Blocklängen wurde eine effiziente Heuristik bestimmt.

Ein ausführlicher Vergleich der Fehlererkennungseigenschaften von bekannten Codes und den neu bestimmten Codes wurde für Retention- und Program-Interference-Fehler in multi-level NAND-Flash-Speicherzellen durchgeführt. Es wurde gezeigt, dass ein Fehlergraph für unidirektionale Fehler mit der maximalen Auslenkung $l = 2$ nicht alle Fehler für multi-level Flash-Zellen abdeckt. Es erwies sich, dass das in der Literatur verwendete Fehlermodell von unidirektionalen Fehlern mit beschränkter Auslenkung nur einen Teil der häufig auftretenden Retention- und Program-Interference-Fehler nicht adäquat modelliert. Experimentell konnte gezeigt werden, dass die in der Arbeit verwendeten Codes zur Erkennung von unidirektionalen Fehlern auch zur Erkennung von Retention- und Program-Interference-Fehlern geeignet sind.

Literaturverzeichnis

[Bo98] Bossert, Martin: Kanalkodierung. B.G. Teubner Stuttgart, 1998. 2. erw. Auflage.

[BP82] Bose, B.; Pradhan, D.K.: Optimal Unidirectional Error Detecting/Correcting Codes. IEEE Transactions on Computers, C-31(6):564–568, June 1982.

[Ca10] Cassuto, Y.; Schwartz, M.; Bohossian, V.; Bruck, J.: Codes for Asymmetric Limited-Magnitude Errors With Application to Multilevel Flash Memories. IEEE Transactions on Information Theory, 56(4):1582–1595, April 2010.

[Ca12] Cai, Yu; Haratsch, Erich F; Mutlu, Onur; Mai, Ken: Error patterns in MLC NAND flash memory: Measurement, characterization, and analysis. In: Design, Automation & Test in Europe Conference & Exhibition (DATE), 2012. IEEE, S. 521–526, 2012.

[EBE13] Elarief, Noha; Bose, Bella; Elmougy, Samir: Limited Magnitude Error Detecting Codes over Z_q. IEEE Transactions on Computers, 62(5):984–989, 2013.

[Fo66] Forney, David G.: Concatenated Codes. Technical report 440, MIT Press, 1966.

[KBE11] Klove, T.; Bose, B.; Elarief, N.: Systematic, Single Limited Magnitude Error Correcting Codes for Flash Memories. IEEE Transactions on Information Theory, 57(7):4477–4487, July 2011.

[MK12] Manzor, Yifat; Keren, Osnat: Amalgamated q-ary codes for multi-level flash memories. In: IEEE Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT) 2012. IEEE, S. 98–103, 2012.

[NKG15] Nieß, G.; Kern, T.; Gössel, M.: Error Detection Codes for Arbitrary Errors Modeled by Error Graphs. In: 11th Workshop on Dependability and Fault Tolerance - VERFE 2015. Porto, Portugal, March 2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft f¨ur Informatik, Bonn 2015

Softwarebasierte Fehlertolerenz für Flash-Speicher von mikrocontroller-basierten Systemen

Patryk Skoncej[1] Felix Mühlbauer[2] Mario Schölzel[3]

Abstract: In diesem Papier wird eine softwarebasierte Methode für die Handhabung von permanenten Fehlern in kostengünstigen Speichern von eingebetteten Systemen vorgestellt. Im Wesentlichen wird der Programmcode derart angepasst, dass keine fehlerhaften Speicherzellen mehr benutzt werden. So ist es möglich permanente Fehler in einfachen und günstigen Speicherbausteinen ohne zusätzlichen Hardwareaufwand zu tolerieren. Dies ist insbesondere für langlebige kostengünstige eingebettete Systeme, z. B. Sensorknoten, interessant, die solche Speicher einsetzen. Verschiedene Methoden für die Handhabung von Speicherfehlern im Daten- und Programmbereich werden betrachtet. Die Methoden für den Programmbereich werden für Flash-Speicher evaluiert. Basierend auf echten Testdaten aus der Produktion von Flash-Speichern wird gezeigt, dass mit der softwarebasierten Methode ein signifikanter Hardwareanteil eingespart werden kann, ohne die Fehlertoleranz von Speicherfehler zu verlieren.

Keywords: Fehlertolerante Systeme, Flash-Speicher, Softwarebasierte Selbstreparatur, Yield-Improvement

1 Einleitung

Die immer weiter schrumpfende Strukturgröße von CMOS-Bausteinen führt zu einer steigenden Verbreitung von eingebetteten elektronischen Geräten, weil mehr Funktionalität bei geringeren Kosten angeboten werden kann. Dies erlaubt die Herstellung von komplexen eingebetteten Systemen mit leistungsstarken CPUs, wie sie zum Beispiel in Steueranlagen von Antriebssystemen im Automobil- und Luftfahrtsektor eingesetzt werden. In solchen Systemen werden etablierte und kostspielige Fehlertoleranzmethoden, wie DMR oder TMR (dual/triple modular redundancy), eingesetzt. Andererseits steigt der Bedarf an kostengünstigen eingebetteten Systemen für das Internet der Dinge und die Industrie 4.0 Infrastruktur. Mit Funkkommunikation ausgestattete Mikrocontroller befinden sich in vielen SmartHome-Anwendungen, wie z. B. Leuchten, Thermostaten und anderen Geräten, die ein Netzwerk bilden, um ihre Umgebung zu Erfassen und zu Steuern. Solche Technologien werden auch eingesetzt, um Produktionsanlagen zu Überwachen und zu Steuern oder für die lebenslange Überwachung von Geräten. Dazu müssen zwei gegensätzliche Bedingungen erfüllt werden: Einerseits müssen die eingesetzten eingebetteten Systeme so kostengünstig wie möglich sein. Zum Beispiel werden Leuchtmittel mit Elektronik entsorgt, sobald sie kaputt gehen. Die eingebaute Elektronik muss folglich kostengünstig sein. Andererseits sollte die eingebaute Elektronik nicht die Hauptursache von Fehlern

[1] IHP Frankfurt (Oder) und BTU Cottbus-Senftenberg, skoncej@ihp-microelectronics.com
[2] Universität Potsdam, muehlbauer@cs.uni-potsdam.de
[3] IHP Frankfurt (Oder) und Universität Potsdam, schoelzel@ihp-microelectronics.com

sein. Zum Beispiel sollte ein Leuchtmittel nicht deshalb kaputt gehen, weil der Mikrocontroller für die Funkkommunikation versagt. Leider steigt die Wahrscheinlichkeit von permanenten Fehlern in CMOS-Bausteinen mit der Verkleinerung der Strukturen [CBT09]. Folglich müssen eingebettete Systeme, so bald sie vermehrt für das Internet der Dinge und Industrie 4.0 eingesetzt werden, eine lange Lebenszeit erreichen und einfache und kostengünstige Komponenten haben. Dafür werden entsprechende Redundanztechniken benötigt. Gewöhnlich wird dies nicht durch das Hinzufügen von Hardwareredundanz (DMR, TMR) erreicht. Eine Lösung könnten softwarebasierte Verfahren sein, indem ein Teil der inhärent verfügbaren Redundanz in diesen eingebetteten Systemen genutzt wird, ohne Extrahardware zu benötigen. Die kleinen eingebetteten Systeme, die in diesem Papier betrachtet werden, sind typischerweise aus sehr einfachen 8- oder 16-Bit Mikrocontrollern mit etwas Speicher und Peripherie aufgebaut. Die Speicher sind eine Mischung aus ROM, RAM und Flash-Speicher mit manchmal nicht mehr als insgesamt 64 kByte. Dennoch belegen die Speicher einen signifikanten Bereich auf dem Chip verglichen mit der CPU. Abbildung 1 zeigt diese Verteilung für einen Mikrocontroller aus der MSP430-Familie mit 64 kByte Flash und 16 kByte RAM.

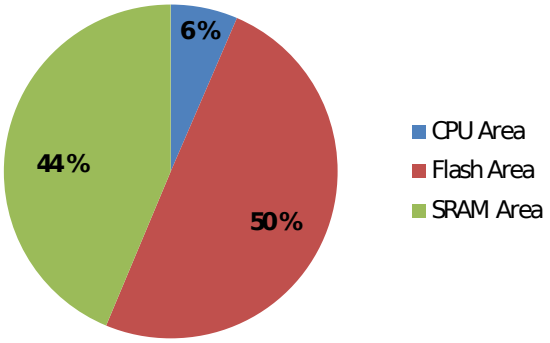


Abb. 1: Belegung der Chipfläche

95 % der Chipfläche wird für Speicher benutzt. Folglich würde eine fehlertolerante CPU die Zuverlässigkeit des Systems nicht signifikant erhöhen, unter der Annahme einer gleich verteilten Fehlerrate. Statt dessen müssen Fehler in den Speichern behandelt werden. Typischerweise wird nicht der komplette Speicher von einer Anwendung verwendet und es verbleiben ungenutzte Speicherwörter. Diese Redundanz wird von den bekannten hardwarebasierten Methoden für permanente Fehler nicht ausgenutzt. Dieses Papier stellt eine softwarebasierte Methode für Prozessoren vor, die sich diese inhärente Redundanz zu Nutze macht. Die Anwendung wird derart im Speicher neu verteilt, dass keine fehlerhaften Speicherzellen mehr benutzt werden. Dies ist möglich, weil die Speicherhierarchie von solchen Systemen sehr einfach ist und z. B. keine Caches oder virtuelle Adressierung zum Einsatz kommt. Die Effizienz dieser Strategie wird für verschiedenen Benchmark-Programme unter Verwendung echter Fehlerdaten aus der Produktion von Flash-Speichern evaluiert.

2 Verwandte Arbeiten

Wegen der sehr regelmäßigen Struktur von Halbleiterspeichern wird Fehlertoleranz typischerweise durch hardwarebasierte Methoden realisiert. Die beiden bekanntesten Ansätze sind: 1) Die Maskierung von Fehlern durch Fehlerkorrigierende Blockcodes (ECC) und 2) Aktive Redundanztechniken die defekte Zellen austauschen.

Das Hauptziel von ECC-basierten Methoden ist die Korrektur von transienten und/oder strahlungsverursachten Speicherfehlern. Die bekanntesten Klassen von Codes sind erweiterte Hamming/Hsiao Codes für Einzelbitfehler und Reed-Solomon (RS) und Bose-Chaudhuri-Hocquengham (BCH) Codes für Mehrbitfehler [SS11]. Mehrbit-ECC Codes erhöhen oft Chipfläche, Laufzeit und Energiekosten, was in kleinen eingebetteten Anwendungen schwer zu akzeptieren ist [AIB10, Ch12]. Deshalb werden für eingebettete Speicher, mit Beschränkungen bezüglich erlaubter Latenzzeit und Chipfläche, bevorzugt SECDED Codes (single error correction and double error detection), wie der erweiterte Hamming/Hsiao Code, eingesetzt. Während ECC-basierte Methoden hauptsächlich für die Maskierung von temporären Fehlern Verwendung finden, zielen aktive Redundanztechniken auf permanente Fehler ab und dienen hauptsächlich der Erhöhung der Fertigungsausbeute. Die geläufigste Implementierung von aktiver Redundanz in eingebetteten Speichern ist das BISR (build-in self-repair) Schema, welches sehr ähnlich für RAM [Li03] und Flash-Speicher [HCW06, HCW10] ist. Die Redundanz kann grundsätzlich in drei Klassen eingeteilt werden: Reserveblöcke, Reservezeilen und/oder Reservespalten. Die Reserven werden von einem BISR-Design verwaltet, welches aus drei Funktionen besteht: 1) Eine BISR-Funktion für die Erkennung von fehlerhaften Speicherelementen 2) Die BIRA (build-in redundancy analysis) bestimmt eine passende Reserve 3) Das AR (address reconfiguration) Modul wird benötigt, um Speicherzugriffe entsprechend zu den Reserveelementen umzuleiten [HCW10, HCW07]. Die Komplexität des BISR-Designs hängt hauptsächlich von der BIRA-Funktion ab, die versucht Reserveelemente effizient zu vergeben. Sowohl die Hardwarekomplexität als auch die Laufzeit der BIRA-Funktion beeinflussen die für eine Reparatur benötigte Art und Anzahl von Reserveelementen. Dabei kann ein blockbasiertes Ersetzen von defekten Speicherbereichen effektiver sein als eine zeilen- oder spaltenbasierte Reparatur. Der Nachteil kann allerdings ein höherer Flächenbedarf auf dem Chip und größere Zuordnungstabellen sein [HCW10, Pe08]. Allein mit der Zeilen- oder der Spaltenredundanz (1D) kann kaum eine hohe Reparaturgeschwindigkeit erreicht werden. Dagegen können durch die Kombination der beiden (2D) sehr hohe Geschwindigkeiten erreicht werden, jedoch ist die 2D-Analyse NP-vollständig, was komplexe BIRA-Algorithmen zur Folge hat [HCW06, HCW07]. Folglich sind die kosteneffizientesten Lösungen für die Reparatur von permanenten Fehlern in kleinen eingebetteten Speicher entweder blockbasierte oder eindimensionale Reparaturverfahren.

Allerdings wurden bisher kaum softwarebasierte Reparaturverfahren für Speicher untersucht. Eine Möglichkeit ist es Fehlertoleranz-Methoden für Hardware in Software zu Implementieren, die typischerweise auf temporäre Fehler abzielen [Li06, Re99]. Solche Ansätze basieren auf der redundanten Ausführung von Befehlen und die Verdoppelung von Daten auf Softwareebene. Durch die Vervielfältigung von Daten können auch permanente Fehler in Datenspeichern erkannt werden. Allerdings wird so auch die benötigte Speicher-

größe verdoppelt und Fehler im Programmspeicher werden nicht berücksichtigt. Um diese Methode auf den Programmspeicher anzuwenden, muss das Programm verdoppelt werden oder der Kontrollfluss muss per Software überprüft werden [Go03, OSM00]. Cheng und Gupta beschreiben eine softwarebasierte Idee für die Handhabung von permanenten Fehlern in der fehlerhafte Speicherzellen vermieden werden [CG11]. Das vorgestellte Verfahren ist sehr ähnlich zu dem in [Sc10] für Prozessoren. Allerdings wird kein Beweis angeboten, dass echte Speicherfehler repariert werden können. Darüber hinaus wird keine Methode zur Vermeidung von fehlerhaften Speicherzellen diskutiert. Beide Aspekte werden im vorliegenden Papier untersucht.

3 Softwarebasierte Fehlerbehandlung in Speichern

Der Speicher wird im Folgenden aus einer funktionalen Sicht, d. h. als ein Vektor von Datenwörtern, betrachtet. Ein physikalischer Defekt kann eines, mehrere oder alle Datenwörter beeinträchtigen und in einem Wort wiederum, eine beliebige Anzahl von Bits. Das grundlegende Konzept für eine softwarebasierte Selbstreparatur von Speichern basiert auf der Methode für Prozessoren in [Sc10, Sc11]. Die inhärent verfügbare Redundanz in den Datenwörtern wird per Software verwaltet. Das bedeutet, dass die Verwendung fehlerhafter Datenwörter vermieden wird, in dem die Anwendung im Programmspeicher an den aktuellen Fehlerzustand des Speichers angepasst wird. Das Programm wird nach dem Herstellungsprozess angepasst, sobald der Flash-Speicher programmiert werden soll. Für die Anpassung wird eine Fehlertabelle des Speichers benötigt. Diese Tabelle enthält die Adressen aller fehlerhaften Datenwörter, die nicht mehr verwendet werden sollen. Ein Speichertest, der nach der Produktion ausgeführt wurde, könnte solche Daten liefern. Die Anwendung wird basierend auf der Tabelle an den Fehlerzustand des Speichers angepasst. Dabei muss zwischen Fehlern die das Codesegment betreffen von Fehlern die das Datensegment betreffen unterschieden werden. Beide Fälle werden in den folgenden Abschnitten untersucht.

3.1 Fehlerbehandlung in Datenspeichern

Das Datensegment wird im RAM abgelegt. Viele kleine eingebettete Systeme sind nur mit ein paar Kilobytes RAM ausgestattet. Dieser Speicher wird benutzt für

- die Speicherung von globale Daten,
- die Organisation der dynamischen Speicherverwaltung (heap) und
- die Implementierung des Stapelspeichers (stack).

Für jeden dieser drei Fälle wird eine andere Vorgehensweise im Umgang mit defekten Datenwörtern benötigt. Sie werden in den folgenden Abschnitten beschrieben.

3.1.1 Globale Daten

Im Bereich für globale Daten werden globale Variablen und Konstanten für die Laufzeit der Anwendung gespeichert. Die Adressen werden während der Verlinkung der Anwendung zugewiesen und der Zugriff auf diese Elemente erfolgt typischerweise über absolute Adressen. Ist ein Element von einem Speicherfehler betroffen, muss es an eine andere Speicheradresse verschoben werden. Dies erfordert eine Anpassung aller zugehörigen Speicherzugriffe im Programmcode. Basierend auf dem Binärcode der Anwendung ist dies sehr schwierig, weil eine differenzierte Datenflussanalyse erforderlich ist, um alle Zugriffe auf den globalen Speicher zu identifizieren. Einfacher ist es die Daten des Linkers zu benutzen. Dort sind alle globalen Symbole und Referenzen auf diese verfügbar. Alle notwendigen Informationen können aus den Linker-Dateien ausgelesen werden und damit ist eine Verschiebung von globalen Symbolen durchführbar.

3.1.2 Dynamische Speicherverwaltung

Auf dem Heap werden dynamisch angeforderte Daten gespeichert. Die dynamische Speicherverwaltung muss entweder vom Betriebssystem oder einer eigenen malloc-Funktion implementiert werden. In beiden Fällen muss die Fehlertabelle des Speichers zur Laufzeit der Anwendung zur Verfügung stehen. Mit Hilfe der Tabelle können fehlerhafte Speicherbereiche bei einer Speicheranfrage übersprungen werden.

Bei einer selbst-implementierten malloc-Funktion kann die Fehlertabelle direkt benutzt werden und in die Speicherverwaltung einfließen. Wenn ein Betriebssystem zum Einsatz kommt kann die malloc-Funktion möglicherweise nicht verändert werden. In diesem Fall, kann die Anwendung jedoch kleine Speicherblöcke anfordern bis der Speicher voll ist, um anschließend alle fehlerfreien Blöcke, und nur diese, wieder freizugeben.

3.1.3 Stapelspeicher

Der Stapelspeicher wird benutzt um lokale Variablen, Parameter und Rückgabewerte von Funktionsaufrufen zu Speichern. Für eingebettete Systeme kann die benötigte Größe des Stapelspeichers oft sehr genau abgeschätzt werden, weil in den meisten Fällen keine rekursiven Funktionen eingesetzt werden. Durch eine Analyse des Aufrufgraphen der Anwendung kann, basierend auf den Rahmengrößen der Funktionen, die maximale Größe des Stapelspeichers bestimmt werden. Die Rahmengröße einer Funktion ist dabei der Bedarf an Stapelspeicher für ihren Aufruf. Mit diesem Wissen kann ermittelt werden, ob der Stapelspeicher fehlerhafte Speicherwörter enthält. Zum Beispiel ist es sehr gewöhnlich den Stapelspeicher an der höchsten Adresse des RAMs beginnen zu lassen. Der Stapelspeicher wächst dann in Richtung kleinerer Adressen. Wenn dieser Speicherbereich fehlerhaft ist, kann der ganze Stapelspeicher in einen andern zusammenhängenden fehlerfreien Speicherbereich verschoben werden. Die Verschiebung ist sehr einfach, weil in eingebetteten Systemen der Stapelzeiger von der Anwendung selbst initialisiert wird und hier eine ande-

re Startadresse genutzt werden kann. Jedoch benötigt diese Methode einen ausreichend großen, zusammenhängenden und fehlerfreien Speicherbereich. Ferner ist möglicherweise auch eine Verschiebung der globalen Daten oder des Heapspeichers notwendig. Zum Beispiel können globale Variablen vor und nach dem Stapelspeicher angelegt werden, um den verfügbaren Speicher effizient zu nutzen.

Für fragmentierte Stapelspeicher dagegen, ist es möglich Prolog und Epilog der Funktionen anzupassen. Im Prolog wird sofort der für die Funktion benötigte Speicher reserviert. Typischerweise geschieht das durch eine Verringerung des Stapelzeigers um eine Konstante. Ist dieser Speicherbereich fehlerhaft, kann der Stapelzeiger entsprechend weiter verringert werden, bis das nicht mehr der Fall ist. Im Epilog wird der belegte Speicherbereich wieder freigegeben. Diese Modifikation erfordert eine Anpassung des Compilers und die Fehlertabelle muss für die dynamische Überprüfung im Prolog zur Laufzeit zur Verfügung stehen.

3.1.4 Zusammenfassung

Mit der vorgestellten Methode ist es möglich Fehler im RAM von eingebetteten Systemen zu handhaben. Für die dynamische Speicherverwaltung und für die effiziente Nutzung eines Speicherbereichs als Stapelspeicher ist eine Fehlertabelle des Speichers zur Laufzeit notwendig. Die vorgestellten Methoden setzen jedoch eine Anpassung der Toolchain (Compiler und Linker) voraus.

Im nächsten Abschnitt wird gezeigt, das für den Programmbereich keine Anpassung der Toolchain notwendig ist.

3.2 Fehlerbehandlung im Programmspeicher

Der Programmcode ist typischerweise im ROM oder in einem Flash-Speicher. Im folgenden werden nur Flash-Speicher betrachtet, weil der Code in ROMs nicht verändert werden kann.

In der Regel wird für eine Anwendung in kleinen eingebetteten Systemen nur ein Binary erzeugt, welches alle Funktionen nacheinander enthält (siehe Abbildung 2 a, Funktion F1 bis F4).

Um die Benutzung von fehlerhaften Speicherwörtern für Programmcode zu vermeiden, werden zwei Wege vorgeschlagen:

- Fehlerhafte Speicherwörter innerhalb einer Funktion sind erlaubt
- Fehlerhafte Speicherwörter sind nur außerhalb von Funktionen erlaubt

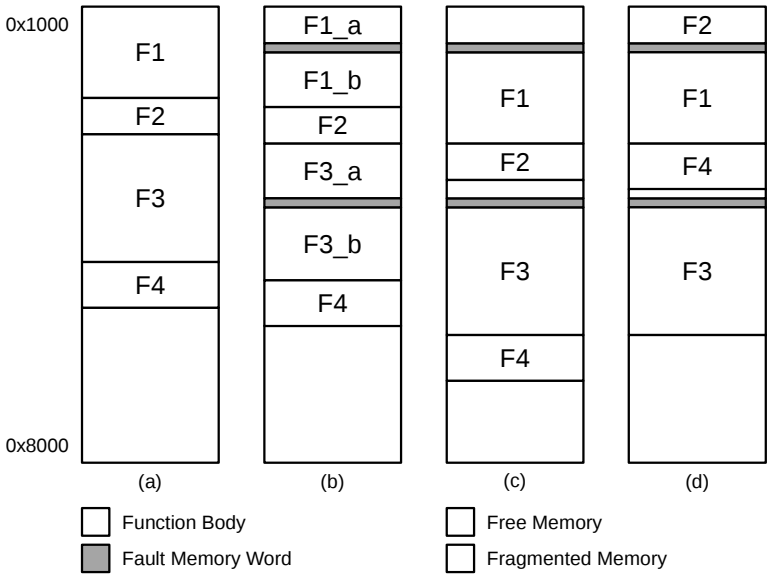


Abb. 2: Fehlerbehandlung im Programmspeicher: (a) Nutzung eines fehlerfreien Speichers (b) Fehlerstellen werden übersprungen (c) Funktionen werden in gleicher Reihenfolge neu platziert (d) Funktionen werden neu verteilt.

3.2.1 Methode 1: Zerteilung von Funktionen

Abbildung 2 b zeigt ein Beispiel für fehlerhafte Speicherwörter innerhalb einer Funktion. Eine Fehlerstelle wird hier durch einen unbedingten Sprungbefehl, der unmittelbar davor eingebaut wird, übersprungen. Auf diese Weise ist eine sehr feingranulare Fehlerbehandlung möglich und es wird nur wenig Speicher verschwendet. Auf der anderen Seite werden Funktionen zerteilt, wie z. B. F1 und F3. Dies erfordert eine Anpassung der verschiedenen Branch-Befehle innerhalb der Funktion. Zum Beispiel verwenden viele Prozessorarchitekturen eine PC-relative Adressierung in bedingten Sprungbefehlen. Sobald ein Sprungbefehl ein fehlerhaftes Speicherwort überspringen würde, muss dieser Befehl angepasst werden. In [MSV13] wurde gezeigt, dass dies sogar auf der Grundlage des Programmcodes möglich, jedoch sehr rechenintensiv, ist.

3.2.2 Methode 2: Neuverteilung von Funktionen

Eine andere Möglichkeit ist es die Funktionen neu anzuordnen ohne deren Implementierung zu verändern, d. h. Leerstellen können sich nur zwischen den Funktionen befinden. Dafür gibt es zwei Möglichkeiten:

- Die Reihenfolge der Funktionen bleibt erhalten (Abbildung 2 c) oder
- die Reihenfolge der Funktionen darf verändert werden (Abbildung 2 d).

Im zweiten Fall kann das Ausmaß an unbenutzten Speicherfragmenten minimiert werden, indem jene Funktionen ausgewählt werden die am besten in eine der Lücken zwischen zwei Fehlerwörtern passen. Dies ist das bekannte Behälterproblem (bin packing problem). Es ist NP-schwer und kann jedoch z. B. mit einfachen heuristischen Methoden gelöst werden.

Ein einfacherer Algorithmus kann benutzt werden, indem die Reihenfolge der Funktionen beibehalten wird. Dies kann jedoch eine starke Fragmentierung nach sich ziehen, wie in Abbildung 2 c dargestellt. Mit dieser Methode müssen nur die Befehle für Funktionsaufrufe angepasst werden, was die Anpassung des Programms vereinfacht.

4 Ergebnisse

In diesem Abschnitt wird das Potential der vorgestellten softwarebasierten Fehlerbehandlung an Hand von echten Produktionsfehlern in Flash-Speichern evaluiert.

4.1 Produktionsausbeute bei Flash-Speichern

Die Evaluation basiert auf echten Testdaten von Flash-Speichern in 250 nm-Technologie. Zwei Versionen werden betrachtet: 8K und 16K. Die 8K-Speicher haben 8192 32-Bit Datenwörter und damit 32 kByte und die 16K-Version 16384 Wörter und damit 64 kByte. Es stehen die Testergebnisse von je 244 Speichern der beiden Varianten, verteilt auf vier Wavern, zur Verfügung. Sie sind in Tabelle 1 zusammengefasst.

Speicherart	Anzahl	Fehlerfrei	Fehlerhaft	Ausbeute
8K	244	210	34	86,1 %
16K	244	195	49	79,9 %

Tab. 1: Produktionsausbeute für 8K- und 16K-Flash-Speicher

Die 8K-Version hat mit 86,1 % eine höhere Ausbeute als die 16K-Variante mit nur 79,9 %. Eine detaillierte Diagnose zeigte, dass in einem Speicher meistens nur wenige Datenwörter defekt sind. In Abbildung 3 ist jeweils die Anzahl der Speicher mit einer bestimmten Anzahl an Fehlerwörtern dargestellt.

Nur bei 14 der 34 defekten Speicher sind alle Datenwörter beschädigt. Ein Speicher ist größtenteils unbrauchbar. In solchen Bausteinen sind typischerweise die Daten- und/oder Adressleitungen fehlerhaft, d. h. nur 6 % der 244 Bausteine habe Defekte die sie unbrauchbar machen. Bei 10 % (19 Stück) sind weniger als 446 Datenwörter beschädigt, d. h. nur weniger als 5,5 % des Speichers kann nicht mehr benutzt werden. Diese Speicher können für Anwendungen eingesetzt werden, die diese kleine Speichereinbusen hinnehmen und die Benutzung der fehlerhaften Speicherwörter vermieden werden kann.

Eine ähnliche Betrachtung wurde für die 16K-Speicher durchgeführt (siehe Abbildung 4). Bei 16 der 49 Bausteinen sind knapp 25 % oder mehr der Speicherwörter betroffen. Andererseits sind in den übrigen 33 beschädigten Speichern nur weniger als 1 % der Wörter

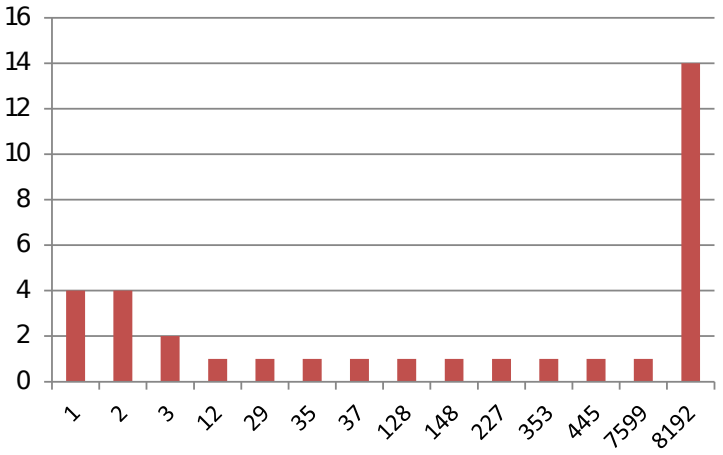


Abb. 3: Verteilung der Fehlerhäufigkeit für 8K-Speicher

fehlerhaft. Auch diese Speicher sind potentielle Kandidaten für Anwendung mit Fehlerbehandlung.

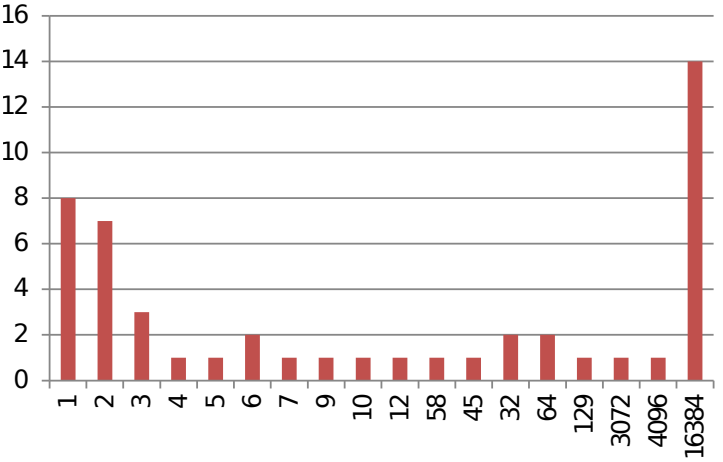


Abb. 4: Verteilung der Fehlerhäufigkeit für 16K-Speicher

4.2 Verbesserung der Produktionsausbeute

Im Folgenden wird dargestellt in wie fern die Produktionsausbeute für die oben beschriebenen Flash-Speicher unter Anwendung der vorgestellten Fehlerbehandlungsmethoden verbessert werden kann. Bei der softwarebasierten Reparatur ist die erreichbare Verbesserung vom Grad der Speichernutzung der Anwendung und der einzelnen Funktionen abhängig. Deshalb werden im Folgenden Diagnoseprogramm vorgestellt, die in Größe und Anzahl ihrer Funktionen variieren (siehe Tabelle 2). Die Speichernutzung liegt zwischen 80–98 %.

Größe der Funktion in Wörtern	Anzahl Funktionen für verschiedene Programme						
	sf80	sf92	lf80	lf90	sn90	sn95	sn98
0–19	109	126	0	0	120	112	108
20–39	183	210	44	48	70	73	77
40–59	18	21	0	0	20	24	21
60–79	0	0	20	24	10	10	13
80–99	0	0	0	0	8	8	8
100–119	0	0	0	0	6	4	2
120–139	0	0	10	12	3	4	6
140–159	0	0	0	0	1	3	2
160–179	0	0	0	0	2	1	1
180–199	0	0	0	0	1	1	2
270–289	0	0	5	6	0	1	1
470–489	0	0	3	3	0	0	0
Anzahl der Funktionen	310	357	82	93	241	241	241
Speichernutzung (Bytes)	6565	7545	6560	7440	7440	7800	8028
Speichernutzung	80 %	92 %	80 %	90 %	90 %	95 %	98 %

Tab. 2: Verschiedene Diagnoseprogramme und deren Aufbau

Die Diagnoseprogramme sf* und lf* wurden künstlich erzeugt und sind aus kleinen bzw. großen Funktionen aufgebaut. Die Funktionen sind darüber hinaus zufällig angeordnet. Die Diagnoseprogramme sn* basieren auf einer echten Anwendung eines Sensorknotens mit einem MSP430-Prozessor. Die Größen der Funktionen wurden dafür geringfügig verändert, um verschiedene Speicherbereiche zu belegen. Die Funktionen wurden außerdem vom Linker in absteigender Größe sortiert.

Für die oben beschriebenen fehlerhaften Speicher werden nun die Diagnoseprogramme mit Hilfe der folgenden Methoden programmiert:

- Methode 1 a: Neuordnung der Funktionen
- Methode 1 b: Neuplatzierung der Funktionen in der gleichen Reihenfolge
- Methode 2: Zerteilung von Funktionen

Für die Neuordnung der Funktionen wurde eine einfache Heuristik verwendet, welche die größte passende Funktion dem nächsten fehlerfreien Speicherbereich zuweist. In Tabelle 3 ist dargestellt, wie viele Anwendungen erfolgreich angepasst werden konnten und welche neue Produktionsausbeute sich insgesamt daraus ergibt.

Die größte Ausbeute wird mit Methode 2 erzielt. Hier wird vor jedem fehlerhaften Speicherwort eine Branch-Operation eingefügt. Folglich muss der Reservespeicher höchstens doppelt so groß sein, wie die fehlerhaften Bereiche. Andererseits müssen auch Sprünge innerhalb der Funktionen angepasst werden. Außerdem wird die Laufzeit der Funktionen verlängert, was bei den anderen Fehlerbehandlungsmethoden nicht der Fall ist. Daher

	Methode 1 a			Methode 1 b			Methode 2		
	Ok	Fehler	Ausbeute	Ok	Fehler	Ausbeute	Ok	Fehler	Ausbeute
sf80	17	2	93,0 %	15	4	92,2 %	19	0	93,8 %
sf92	16	3	92,6 %	15	4	92,2 %	17	2	93,0 %
lf80	15	4	92,2 %	15	4	92,2 %	19	0	93,8 %
lf90	15	4	92,2 %	15	4	92,2 %	18	1	93,4 %
sn90	15	4	92,2 %	15	4	92,2 %	18	1	93,4 %
sn95	15	4	92,2 %	14	5	91,8 %	16	3	92,6 %
sn98	15	4	92,2 %	11	8	90,5 %	14	5	91,4 %

Tab. 3: Produktionsausbeute insgesamt bei 8K-Speichern für verschiedene Fehlerbehandlungsmethoden und Anwendungen

ist für sn98 Methode 2 schlechter als Methode 1 a. Insgesamt wird selbst mit der sehr einfachen Methode 1 b eine Ausbeute von über 90 % erzielt und damit um mehr als 4 % verbessert.

Ohne Neuordnung der Funktionen sinkt die Ausbeute für Anwendungen mit hohem Speicherverbrauch. Mit Neuordnung ist die Ausbeute quasi unabhängig vom verfügbaren Reservespeicher.

	Methode 1 a			Methode 1 b			Methode 2		
	Ok	Fehler	Ausbeute	Ok	Fehler	Ausbeute	Ok	Fehler	Ausbeute
sf80	33	0	93,4 %	33	0	93,4 %	33	0	93,4 %
sf92	33	0	93,4 %	33	0	93,4 %	33	0	93,4 %
lf80	33	0	93,4 %	33	0	93,4 %	33	0	93,4 %
lf91	33	0	93,4 %	33	0	93,4 %	33	0	93,4 %
sn90	33	0	93,4 %	33	0	93,4 %	33	0	93,4 %
sn95	33	0	93,4 %	33	0	93,4 %	33	0	93,4 %
sn98	33	0	93,4 %	31	2	92,6 %	33	0	93,4 %

Tab. 4: Produktionsausbeute insgesamt bei 16K-Speichern für verschiedene Fehlerbehandlungsmethoden und Anwendungen

Die gleiche Auswertung wurde ebenfalls für die 16K-Speicher durchgeführt. Die Anzahl der Funktionen der einzelnen Diagnoseprogrammen in Tabelle 2 wurde dazu verdoppelt. Die Ergebnisse sind in Tabelle 4 zusammengefasst. Methode 1 a und 2 können alle Defekte reparieren. Selbst die einfachste Methode 1 b kann fast alle Fehler behandeln; außer bei der Anwendung mit 98 % Speichernutzung.

In Tabelle 5 ist die erreichbare Verbesserung der Produktionsausbeute für 8K- und 16K-Bausteine zusammengefasst.

Für 8K-Bausteine kann die Ausbeute, je nach Speicherauslastung der Anwendung und der eingesetzten Reparaturmethode, von 86 % auf etwa 92 % erhöht werden, bei den 16K-Bausteinen sogar von 80 % auf etwa 93 %. Dabei wird die Ausbeute verbessert ohne zusätzliche Reservespeicher oder einer Modifikation des Speichercontroller. Folglich kann es günstig sein, die vorgestellten Methoden für einfache eingebettete Speicher anzuwen-

Speicherart	Anzahl	Ausbeute			
		ohne Reparatur	Methode 1 a	Methode 1 b	Methode 2
8K	244	86,1 %	92–93 %	90–92 %	91–94 %
16K	244	79,9 %	93 %	92–93 %	93 %

Tab. 5: Verbesserung der Produktionsausbeute für 8K- und 16K-Flash-Speicher mit softwarebasierter Reparatur

den, bei denen diese Hardwareerweiterungen aus Kostengründen nicht möglich sind. Es ist jedoch ein weiterer Schritt im Produktionsprozess erforderlich, der die Anpassung der Anwendung an Hand der gewählten Reparaturmethode durchführt.

5 Zusammenfassung

Es wurden verschiedene softwarebasierte Methoden vorgestellt, um permanente Fehler in Daten- und Programmspeichern in kleinen eingebetteten System zu Handhaben. Basierend auf detaillierten Diagnosedaten von produzierten Speicherbausteinen konnte gezeigt werden, dass softwarebasierte Methoden es ermöglichen mehr als 50 % der Fehler zu reparieren. Dabei kann im Gegensatz zu den hardwarebasierten Methoden ohne zusätzliche Hardware ausgekommen werden, indem die Redundanz in Form von unbelegten Speicherwörtern ausgenützt wird. Die vorgestellten Methoden eignen sich daher sehr gut die Produktionsausbeute von kleinen eingebetteten System zu Verbessern, die aus einfachen und kostengünstigen Komponenten aufgebaut sind, wie z. B. Flash-Speicher ohne Selbstreparatur.

Literaturverzeichnis

[AIB10] Ankolekar, P.; Isaac, R.; Bredow, J.: Multibit error-correction methods for latency-constrained flash memory systems. IEEE Transactions on Device and Materials Reliability, 10(1):6, 2010.

[CBT09] Cao, Yu; Bose, Pradip; Tschanz, Jim: Reliability Changes in Nano-CMOS Design (Guest Editor's Introduction). IEEE Design & Test of Computers, 26(6):6–7, 2009.

[CG11] Cheng, Da; Gupta, Sandeep: A novel software-based defect-tolerance approach for application-specific embedded systems. In: IEEE 29th International Conference on Computer Design (ICCD'11). S. 443–444, 2011.

[Ch12] Chu, Chia-Ching; Lin, Yi-Min; Yang, Chi-Heng; Chang, Hsie-Chia: , A Fully Parallel BCH Codec with Double Error Correcting Capability for NOR Flash Applications, 2012.

[Go03] Goloubeva, O.; Rebaudengo, M.; Reorda, M.S.; Violante, M.: Soft-error detection using control flow assertions. In: Defect and Fault Tolerance in VLSI Systems, 2003. Proceedings. 18th IEEE International Symposium on. S. 581–588, Nov 2003.

[HCW06] Hsiao, Yu-Ying; Chen, Chao-Hsun; Wu, Cheng-Wen: , A built-in self-repair scheme for NOR-type flash memory, 2006.

[HCW07] Huang, Rei-Fu; Chen, Chao-Hsun; Wu, Cheng-Wen: Economic Aspects of Memory Built-in Self-Repair. IEEE Design & Test of Computers, 24(2):8, 2007.

[HCW10] Hsiao, Yu-Ying; Chen, Chao-Hsun; Wu, Cheng-Wen: Built-In Self-Repair Schemes for Flash Memories. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 29(8):13, 2010.

[Li03] Li, Jin-Fu; Yeh, Jen-Chieh; Huang, Rei-Fu; Wu, Cheng-Wen: A built-in self-repair scheme for semiconductor memories with 2-D redundancy. In: Test Conference, 2003. Proceedings. ITC 2003. International. Jgg. 1, S. 393–402, 2003.

[Li06] Lisboa, C.A.L.; Carro, L.; Reorda, M.S.; Violante, M.: Online hardening of programs against SEUs and SETs. In: Defect and Fault Tolerance in VLSI Systems, 2006. DFT '06. 21st IEEE International Symposium on. S. 280–290, Oct 2006.

[MSV13] Müller, Sebastian; Schölzel, Mario; Vierhaus, Heinrich Theodor: Towards a Graceful Degradable Multicore-System by Hierarchical Handling of Hard Errors. In: Parallel, Distributed and Network-Based Processing (PDP), 2013 21st Euromicro International Conference on. S. 302–309, Feb 2013.

[OSM00] Oh, N.; Shirvani, P.P.; McCluskey, E.J.: Control-Flow Checking by Software Signatures. IEEE Transactions on Reliability, 51(2):111–122, 2000.

[Pe08] Pekmestzi, K.; Axelos, N.; Sideris, I.; Moshopoulos, N.: , A BISR Architecture for Embedded Memories, 2008.

[Re99] Rebaudengo, M.; Reorda, M.Sonza; Torchiano, M.; Violante, M.: Soft-error Detection through Software Fault-Tolerance techniques. In: Proc. of the 14th International Symposium on Defect and Fault-Tolerance in VLSI Systems (DFT'99). Proc. of the 14th International Symposium on Defect and Fault-Tolerance in VLSI Systems (DFT'99), S. 210–218, 1999.

[Sc10] Schölzel, Mario: Software-Based Self-Repair of Statically Scheduled Superscalar Data Paths. In: Proc. of the 13th IEEE International Symposium on Design & Diagnostics of Electronic Circuits & Systems (DDECS'10). Proc. of the 13th IEEE International Symposium on Design & Diagnostics of Electronic Circuits & Systems (DDECS'10), S. 66–71, 2010.

[Sc11] Schölzel, Mario: Fine-Grained Software-Based Self-Repair of VLIW Processors. In: 26th IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems. 26th IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems, S. 41–49, 2011.

[SS11] Skoncej, Patryk; Schoof, Gunther: , Getting ready for non-volatile memories, 2011.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Möglichkeiten der Nutzung von RRAM in Low-Power Microcontrollern

Frank Vater [1], Goran Panic [2], Mario Schölzel [3]

Abstract: Die Energieeffizienz aktueller Microcontroller wird durch die aggressive Nutzung von verschiedenen Ruhe- und Schlafmodi erreicht. Eine vollständige Abschaltung des Microcontrollers ist dabei aufwendig, da die Zustände in flüchtigen Speichern nicht beibehalten werden. Nichtflüchtige Speicher bieten die Möglichkeit auf einfach Weise die Daten zwischenzuspeichern. In diesem Paper werden die verschiedenen Speichertechnologien in Hinblick auf ihre Energieeffizenz unter Berücksichtigung von Schreib-/Lesegeschwindigkeit sowie Lebensdauer untersucht. Dabei findet eine sehr genaue und technologienahe Simulation auf Basis von Transistormodellen statt.

Keywords: Microcontroller, Low Power, RRAM, Schattenregister, Registerfile

1 Einleitung

Microcontroller und Mikroprozessoren finden sich vielerorts wieder. Das kann für einfache Steuerungs- und Regelungsaufgaben sein bis hin zu komplexen Prozesssteuerungen. Häufig werden dafür Prozessoren mit 8, 16 oder 32 Bit Datenbreite benutzt. Während es z.B. in der Prozessautomatisierung keinerlei Probleme bereitet den Prozessor permanent mit elektrischer Energie zu versorgen, ist das bei einem Einsatz z.B. in einem Sensornetzwerk auf dem freien Feld als Bodensensor schwieriger. Gerade wenn eine Vielzahl von Geräten ausgebracht wurde, können diese unmöglich in kurzen Zeitabständen eingesammelt und die Batterien ausgetauscht oder Akkus aufgeladen werden. Daher ist es unabdingbar Stromsparmechanismen in solchen Geräten einzusetzen. Obwohl gerade kleinere Prozessoren der Entwicklung nach dem Mooreschen Gesetz hinterherhinken, sind hier nun auch Strukturgrößen von 250nm und deutlich kleiner erreicht. Dabei tritt verstärkt der Effekt der Leckströme auf.

In Microcontrollern werden unterschiedliche Speichertypen eingesetzt. Die Abb. 1 zeigt dabei die Speicherarchitektur eines typischen Microcontrollers. Der Flash dient dabei zum Ablegen des auszuführenden Programms. So kann Prozessor von der Energieversorgung getrennt werden ohne dass das Programm verloren geht. Teure, im Sinne von energieintensiven, Schreibzugriffe sind recht selten und treten nur beim Programmieren des Microcontrollers oder bei einem Code-Update auf. Da im Feld ein Code-Update selten vorkommt, sind vornehmlich die Leseoperation und der Leckstrom aus energetischer Sicht interessant.

[1] IHP GmbH, Systeme, Im Technologiepark 25, 15236 Frankfurt(Oder), vater@ihp-microelectronics.com
[2] IHP GmbH, Systeme, Im Technologiepark 25, 15236 Frankfurt(Oder), panicl@ihp-microelectronics.com
[3] IHP GmbH, Systeme, Im Technologiepark 25, 15236 Frankfurt(Oder), schoelzel@ihp-microelectronics.com

Neben dem Programm- wird auch Datenspeicher benötigt, der als SRAM umgesetzt wird. In diesem werden Variableninhalte abgelegt, die während der Abarbeitung komplexerer Programme benötigt werden und die nicht mehr im Registerfile abgelegt werden können.

Im unteren Speicherbereich befinden sich ein “memory-mapped” IO. An diesen sind die peripheren Komponenten angeschlossen, um somit einen einfachen Zugriff darauf zu ermöglichen. Diese Adressen beherbergen keinen “echten” Speicher, vielmehr sind sie eine Schnittstelle zu anderen Komponenten. Beispielsweise können so GPIOs programmiert oder serielle Schnittstellen für den Datenaustausch bedient werden.

Ein weiterer Speicher im Microcontroller, das Registerfile, ist nicht direkt adressierbar und daher auch nicht in Abb. 1 dargestellt. Das Registerfile ist typischerweise aus Flipflops aufgebaut.

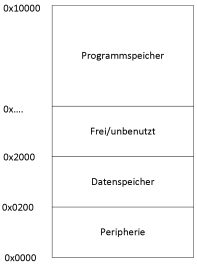


Abb. 1: Speicheraufteilung eines Microcontrollers

Als Beispiel für einen Microcontroller kann der MSP430 der Firma Texas Instruments (TI) dienen [Te15a]. Dieser hat einen 16 Bit breiten Datenpfad und verfügt über einen Adressraum von 64kByte mit getrennten Programm- und Datenspeicher. Aufgrund der niedrigen Taktraten und dem Einsatzgebiet als Sensorknoten ist das Design ohne Cache ausgeführt. Dem Anwendungsgebiet entsprechend sind verschiedene Stromsparmodi implementiert. Diese werden anhand des MSP430 von TI, Familie 5, hier kurz vorgestellt [Te15b]. Neben dem Arbeitsmodus (“active”) stehen die Schlafmodi LPM0 bis LPM4 zur Verfügung. Diese unterscheiden sich untereinander im wesentlichen durch das Abschalten der drei Clock Domänen.

Darüber hinaus existieren noch die Modi LPM 3.5 und 4.5. In diesen Tiefschlaf-Modi verlieren alle flüchtigen Speicher ihren Inhalt. Damit sind RAM und Registerinhalte beim Erwachen aus dem Tiefschlaf gelöscht. Die folgende Startsequenz entspricht damit also dem Ablauf wie bei einem regulären Power-On oder Reset. Messwerte oder Variablen, die nicht im Flash gesichert wurden, sind damit verloren. Ebenso muss somit die Initialisierungsphase wieder durchlaufen werden.

In der Entwurfsphase eines Microcontrollers wird dieser in einer Hardwarebeschreibungssprache wie VHDL implementiert. Daraus wird später eine Netzliste generiert und ein Layout überführt. Der benötigte Speicher wird mit Hilfe eines technologieabhängigen Speichergenerators erzeugt. Dieser fertige Block wird vom Chipdesigner in das Layout eingebunden. Eine Simulation der Leistungsaufnahme des Gesamtsystems, also Digitalzellen, Speicher und I/O Pads, ist sehr zeitaufwendig und kann etliche Tage bis Wochen

in Anspruch nehmen. Darüber hinaus verfügt der Chipdesigner oft nicht über alle notwendigen Bibliotheken (Transistormodelle) der jeweiligen Technologie, die für die Chipfertigung benutzt wird. In einer 0.25 μm Technologie des IHP [IH15] sind sowohl die Speicher als auch die Transistormodelle vorhanden, so dass eine Simulation möglich ist. Die Simulation der einzelnen Speicherblöcke, insbesondere in Hinblick auf die Leistungsaufnahme, kann als Grundlage für eine genauere Bestimmung des Energieverbrauchs eines ASIC genutzt werden. Auf dieser Basis wird eine Exploration der verschiedenen Speichertypen vorgenommen.

2 Related Work

2.1 Stromsparmechanismen

Im Gegensatz zu anderen Halbleitertechnologien ist die CMOS Technologie von Hause aus sehr energieeffizient. Die dynamischen Ströme fließen kurzzeitig bei Schaltvorgängen und die statischen Ströme beschränken sich auf Leckströme. Anfänglich arbeitete die CMOS Technologie bei einer Spannung von 5 V. Mit kleineren Strukturgrößen war es möglich, die Betriebsspannung immer weiter abzusenken, so dass bei Strukturgrößen um 65 nm Spannungen um 1 Volt genügen. Da die Spannung quadratisch in die Leistungsaufnahme eingeht, ist das Einsparpotential bei dynamischen Umschaltvorgängen erheblich. In Designs mit unterschiedlichen Anforderungen an die Verarbeitungsgeschwindigkeit, können die Blöcke einzeln regulierbare Spannungsversorgungen erhalten. So kann z.B. ein langsamer Peripherieblock mit niedrigerer Spannung arbeiten als die Ausführungseinheit der CPU, welche auf schnelle Befehlsabarbeitung optimiert ist.

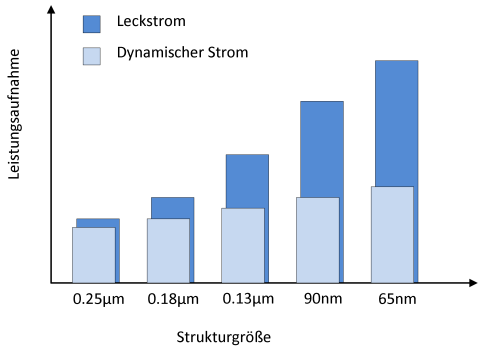


Abb. 2: Dynamischer und statischer Strom in Abhängigkeit von der minimalen Strukturbreite

Gerade der Bereich des dynamischen Schaltens bietet Potential zur Energieeinsparung, welches z.B. durch Clock Gating oder Data Gating umgesetzt werden kann. Mithilfe von Clock Gating können nicht genutzte Partitionen in einem ASIC gezielt von den Umschaltvorgängen ausgeschlossen werden. Da das Clocknetzwerk an jedes Flipflop angeschlossen ist, finden einerseits für das Clocknetzwerk (Treiber) und andererseits in den Flipflops Umschaltvorgänge statt, die Energie verbrauchen. Sollte nun ein Teil eines Design nicht an der Datenverarbeitung beteiligt sein, so wird das Teilnetz der Clock nicht geschaltet. An dieser

Stelle ist die Unterstützung durch die Syntheseprogramme soweit fortgeschritten, dass der Clock Gating Mechanismus fehlerfrei automatisch implementiert wird.

Jedoch steigen im Submicron-Bereich die Leckströme erheblich an. Grund dafür sind die geringeren Strukturgrößen sowie die geringen Thresholdspannungen. Die Abb. 2 zeigt die Entwicklung von dynamischen und statischen Strömen bei kleiner werdenden Strukturgrößen. Vor allem im Einsatzbereich von Sensorknoten, welche über einen langen Zeitraum energieautark fungieren müssen, ist ein hoher Stromverbrauch im Standby nicht praktikabel. In kurzer Zeit wäre das Lebensende bei einem batteriebetrieben Gerät erreicht. Aktuell sind die Strukturgrößen von Microcontrollern für Sensorknoten im Bereich von 180 bis 130 nm, wobei auch hier der Trend aus Kostengründen zu immer kleineren Strukturen geht. Ab etwa 90nm werden die Leckströme so dominant, dass dafür geeignete Mittel gefunden werden müssen, diese auf ein Minimum zu beschränken.

Ein effizientes Mittel gegen hohe Leckströme ist der Einsatz der Power Gating Technik [Pa14]. Bei dieser werden die Blöcke in einem Design separat mit Energie versorgt. Im Gegensatz zum Clock Gating ist beim Power Gating ein aktives Eingreifen durch den Nutzer/Programmierer sinnvoll. Da dieser den Anwendungszweck kennt, kann er aus seiner Position heraus gut abschätzen, welche Blöcke komplett abgeschaltet werden können, so dass der Leckstrom auf Null sinkt.

Der Nachteil des Power Gating begründet sich darin, dass Registerinhalte von flüchtigen Speichern verloren gehen können. Soll ein Microcontroller mit Power Gating in einen sehr energiearmen Schlafmodus versetzt werden, so ist darauf zu achten, dass das Registerfile in einem nichtflüchtigen Speicher gesichert wird. Beim Erwachen, ausgelöst durch einen Interrupt, verweist der Programmzähler auf eine Adresse zur Interruptbehandlung. Innerhalb dieser Routine muss ein gesicherter Zustand aufwendig wiederhergestellt werden. Dieses kostet Zeit und der zusätzliche Aufwand ist nicht erwünscht.

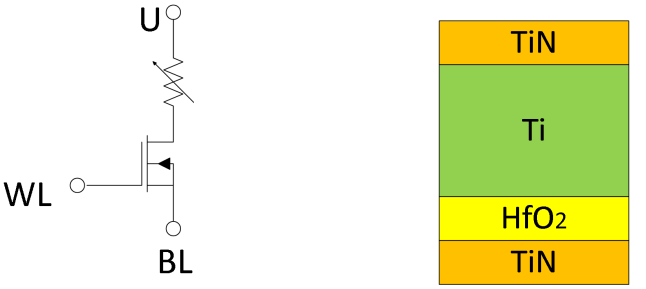
2.2 RRAM

Die Möglichkeit eine Information mit Hilfe eines änderbaren Widerstands zu speichern, ist schon seit den 1960er Jahren bekannt [NB64]. Jedoch erst mit der in [Le08] beschriebenen Lösung war es möglich, diese Art Speicherelement in die siliziumbasierten Chipproduktion zu integrieren.

In Abb. 3a wird der Schaltplan eines Bits in RRAM Technologie gezeigt. Dem Transistorausgang ist ein änderbarer Widerstand nachgeschaltet. Dieser Widerstand kann durch einen kurzen elektrischen Impuls seinen Wert verändern. Die Widerstandsänderung bleibt auch nach Abschalten der Versorgungsspannung erhalten. Auf diese Weise kann entweder eine logische Null oder Eins nichtflüchtig abgespeichert werden. Konstruktiv wird das mittels eines speziell präparierten Vias gelöst. In Abb. 3b ist der vertikale Aufbau eines Transistors mit dem dazu gehörigen modifizierten Via zu sehen. Das Standard-Via, bestehend aus Titan (Ti) und Titannitrit (TiN), befindet sich zwischen Metal2 und Metal3. Das unteren Ende des Vias wird durch das Einbringen von Hafniumoxid (HfO_2) modifiziert, welches so die Funktionalität des veränderbaren Widerstands ermöglicht. Da der RRAM

im Backend, also erst in der Verdrahtungsebene, eingebracht wird, ist diese Technologie vollständig CMOS kompatibel.

Nach der Produktion des Speichers, muss dieser einmalig initialisiert werden. Dazu wird ein Strom von 10μA an jede Speicherzelle angelegt. Da diese Initialisierungsphase auf einem Tester durchgeführt werden kann, wird dieser Energieverbrauch nicht weiter im funktionalen Betrieb des Microcontrollers betrachtet.



(a) Schaltplan einer RRAM Zelle

(b) Querschnitt durch ein Via, dass als RRAM Zelle ausgeführt ist

Abb. 3: RRAM Zelle

3 Anwendungsmöglichkeiten

Unabhängig vom verwendeten Microcontroller, ist der Ablauf der Programmabarbeitung stets ähnlich. Nach dem Power-On oder Reset wird das erste Datenwort von einer Basisadresse abgeholt. Diese Adresse liegt typischerweise innerhalb eines nichtflüchtigen Speichers wie einem Flash. Im weiteren wird das dort abgelegte Programm abgearbeitet, das zunächst aus einem Abschnitt mit Konfigurationsdaten besteht. Dabei werden Timer, GPIOs etc. programmiert. Daran schließt sich die Abarbeitung einer Funktion an, die der Hauptbestandteil des Programms ist. Z.B. können Messwerte aufgenommen, verarbeitet und drahtlos an benachbarte Knoten gesendet werden.

Aus Energiespargründen ist es sinnvoll, Messwerte nicht direkt weiter zu leiten. Gegebenenfalls werden Daten in Registern oder im RAM abgelegt, kumuliert und erst nach einer gewissen Zeit gesammelt weiter versendet. Nach dem Erfassen von Daten wird der Microcontroller in einen Schlafmodus versetzt, um so die Lebensdauer der Batterie zu verlängern. Dieser Energiesparmodus (z.B. LPM 0 bis LPM 4 beim MSP430) wird durch ein Ereignis an einen GPIO Pin oder nach Ablauf eines Timers beendet.

In Abhängigkeit des Anwendungsszenarios kann der Zustand des Schlafmodus für Sekunden, Minuten, Stunden oder noch länger beibehalten werden. Je länger die Schlafzeit ist, umso mehr lohnen sich die Tiefschlafmodi wie der LPM 3.5 und LPM 4.5 beim MSP430. Bei diesen wird die Leistungsaufnahme durch das Abschalten der Spannungsversorgung für Registerfile und RAM sehr weit heruntergefahren. Es können jedoch Fälle eintreten, in denen der Wechsel in diese Modi sehr aufwendig ist. Werden in den Registern oder

im RAM Daten vorgehalten, die beim nächsten Erwachen zur Verfügung stehen sollen, müssen diese in einem nichtflüchtigen Speicher gepuffert werden. Hierbei kann der Einsatz von RRAM die Sicherungsoperation deutlich vereinfachen.

4 Ersetzungsszenarien

Wie in Abschnitt 1 beschrieben, enthält ein Microcontroller verschiedenartige Speichertypen. Da es sich beim RRAM um einen nichtflüchtigen Speicher mit schnellen Zugriffen vor allem beim Lesen handelt, können Überlegungen angestellt werden, alle Speicher durch Speicher diesen Typs zu ersetzen oder zumindest zu ergänzen. Dieser Ansatz kann dann zum Einsatz kommen, wenn der Microcontroller z.B. auf Basis des openMSP430 [Gi09] für ein bestimmtes Projekt gezielt entworfen und gefertigt wird. Sofern seitens der Foundry verschiedene Speichertechnologien angeboten werden, kann der Designer entscheiden, ob z.B. RRAM letztendlich zum Einsatz kommt.

In der Abb. 4 sind die möglichen Szenarien dargestellt, den RRAM in einem Microcontroller einzusetzen. Zunächst ist naheliegend den, als Programmspeicher genutzten, Flash (Abb. 4(a)) durch RRAM (Abb. 4(b)) zu ersetzen. Beispielsweise hat Panasonic bereits den Microcontroller MN1010L im Angebot [Pa15], bei welchem der Programmspeicher in RRAM Technologie ausgeführt ist. Da die notwendige Leistung zum Speichern eines Bits deutlich unter dem des Speicherns in einem Flash liegt und darüber hinaus auch das aufwendige Löschen von Flashsegmenten entfällt, kann das als Fortschritt betrachtet werden.

In der nächsten Stufe kann auch der SRAM durch einen RRAM ersetzt werden (Abb. 4(c)). Dafür sind jedoch neben der Zugriffszeit und der notwendigen Enerige auch die Anzahl der Schreibzyklen entscheidend. In der Literatur wird die Schreibzeit im Nano- bis Mikrosekunden angegeben und die Anzahl der möglichen Schreibzyklen liegt zwischen 10^6 und 10^9. Läuft also der Microcontroller mit einer Taktfrequenz von 10 MHz und es finden in der aktiven Phase nur wenige Schreibzyklen auf die gleiche Adresse im RAM statt, so kann der SRAM durch RRAM ersetzt werden. Allerdings eventuell muss ein höherer Energiebedarf in der aktiven Phase mit in die Betrachtung einfließen.

In letzter Konsequenz kann auch das Registerfile durch einen RRAM ersetzt werden (Abb. 4(d)), um den Bootprozess nach dem Aufwachen aus einem Tiefschlafmodus zu vermeiden. Jedoch sind die Anzahl der Schreibzyklen während der Abarbeitung eines Programms recht hoch, so dass nach kurzer Zeit keine zuverlässige Speicherung von Daten im RRAM mehr möglich ist. Der Programmzähler ist dabei der kritischste Punkt. Bei einem Dutycycle, also einer aktiven Zeit des Microcontrollers, von 0,1% und möglichen 10^9 Schreibzyklen, ergibt sich eine Lebensdauer von etwa 28 Stunden für den Microcontroller. Anschliessend ist der RRAM in seiner Funktion nicht mehr zuverlässig. Daher ist dieser Ansatz den RRAM als Ersatz für ein Registerfile zu integrieren nicht praktikabel.

Alternativ soll daher betrachtet werden, den RRAM als Schattenregister zu verwenden, wie in Abb. 4(e) gezeigt. Dieses ist parallel zum Registerfile in dem Microcontroller eingebaut. Jeder Schreibzugriff auf das Registerfile wird nicht unmittelbar auf das Schat-

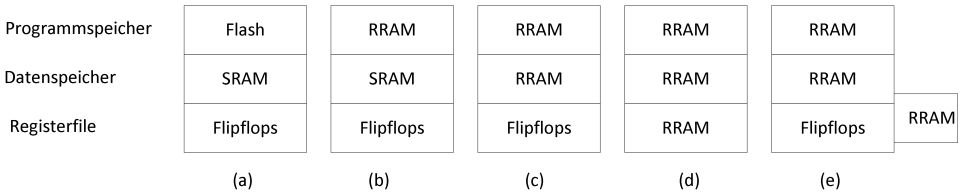


Abb. 4: Ersetzungsstrategien von Speicher in einem Microcontroller durch RRAM

tenregister in RRAM Technologie übertragen. Vielmehr erfolgt dies erst beim Übergang in einen der Schlafmodi, die auch die Spannungsversorgung für SRAM und Registerfile abschalten.

Um auch das Registerfile von der Spannungsversorgung trennen zu können und dennoch die Programmabarbeitung an definierter Stelle fortsetzen zu können, müssen die Registerinhalt also in einem nichtflüchtigen Speicher gesichert werden. Das Ablegen von Werten in nichtflüchtigen Speichern, und ebenso das Zurückschreiben in die Flipflops, benötigt den Einsatz einer gewissen Energiemenge. Bevor also der Prozessor in den Tiefschlafmodus versetzt wird und die Registerinhalte in einen permanenten Speicher überführt werden, muss abgeschätzt werden, ob der Energieaufwand für das Speichern und Wiederherstellen der Werte diesen Schritt rechtfertigt. Das heißt also, dass die Energie die innerhalb des Registerfiles für die voraussichtliche Schlafdauer für Leakage verbraucht wird höher sein muss, als für das Backup. Da diese Werte erheblich von der verwendeten Technologie und der Technik des nichtflüchtigen Speichers abhängen, kann keine pauschale Aussage getroffen werden, ab welcher Schlafdauer sich der Einsatz eines solchen Ansatzes lohnt. Der Nutzer muss also anhand von Programm und den verfügbaren Speichern eine Abschätzung treffen.

Das Schattenregister und der zugehörige Controller können mit Power Gating in das Design integriert werden. So tritt im aktiven Modus keine Belastung der Energiebilanz durch die zusätzliche Komponente auf. Zusätzlich kann das Schattenregister eingesetzt werden, um Checkpoints zu setzen. Damit es möglich, den Programmablauf an einer definierten Stelle fortzusetzen falls z.B. ein Watchdog eine Endlosschleife erkennt und eigentlich ein Reset auslösen müsste.

Die notwendige Energie zum Speichern eines Bitwertes in einer RRAM Zelle liegt aktuell deutlich über der nötigen Energie für ein Flipflop (vgl. Tab. 1). In der Simulation wurde eine um Faktor 10 höhere Energie für das Schreiben eines Wertes und ein Faktor 400 für das Lesen ermittelt. Die Register sind in beinahe jedem Takt des Microcontrollers mit Lese- oder Schreibzugriffen belastet. Während die Anzahl der Lesezugriffe keinen Einfluss auf die Lebensdauer der RRAM Zelle hat, verkürzt jeder Schreibzugriff diese. Dieser Sachverhalt ist bei der Implementierung zu beachten.

5 Simulation der Leistungsaufnahme

Grundsätzlich sind zwei Arten der Leistungsaufnahme zu berücksichtigen, der aktive Modus und der Ruhemodus. Bei Sensorknoten zählt insbesondere der Ruhemodus, der durch Leckströme bestimmt ist, da die Geräte sehr lange in Standby sind, jeweils unterbrochen von kurzen Wachphasen. Allerdings soll auch in den kurzen Wachphasen der dynamische Stromverbrauch möglichst gering sein.

Ein Microcontroller ist in der Anzahl der Transistoren, verglichen mit einem High-End Prozessor, überschaubar. Dennoch ist es nur mit sehr großen Aufwand möglich, diesen Prozessor analog zu simulieren. Eine solche Simulation wäre zwar sehr genau in bezug auf die Leistungsaufnahme, doch ist mit einer Simulationszeit von mehreren Tagen zu rechnen. Für die Bewertung, ab welcher Schlafdauer es sich lohnt auf einen nichtflüchtigen Zwischenspeicher zurückzugreifen, genügt es die Leistungsaufnahme der jeweiligen Speicher zu kennen.

In einer gut charakterisierten Technologie ist eine analoge Simulation des Speichers in Bezug auf Zeitverhalten und Energieverbrauch sehr nahe an den realen Messwerten für die Leistungsaufnahme von dem gefertigten ASIC. Daher kann auf diese Simulationsergebnisse für die Betrachtung des Energieverbrauchs zurückgegriffen werden.

5.1 Digitale Standardzellen

Ein typisches Registerfile kann aus Flipflops aufgebaut werden. Dieses wird zunächst in einer Hardwarebeschreibungssprache beschrieben und anschließend in der Synthese in eine Gatternetzliste überführt. In diesem Schritt ist eine sehr grobe Abschätzung der Leistungsaufnahme möglich. Dabei schätzt das Synthesetool die Anzahl der Umschaltvorgänge von Gattern oder es lässt sich ein Umschaltwert vom Nutzer angeben.

Eine sehr genaue Abschätzung der Leistungsaufnahme ist mit dem Tool “PrimeTime” der Firma Synopsys möglich [Sy15]. Dazu wird zunächst der Prozessor mit dem Programm simuliert, das später im Feld eingesetzt werden soll. Im Gegensatz zu einer üblichen Simulation werden bei dieser Art von Simulation alle Signaländerungen in einer Datenbank gespeichert. Da hierbei einige tausend bis zehntausend Signale und deren Änderungen mit Zeitstempel protokolliert werden müssen, kann die Datenbank eine Größe von mehreren Gigabyte schnell überschreiten. Diese Datenbank kann in das Programm “PrimeTime” eingelesen werden. Dazu wird noch eine technologieabhängige Datenbank eingelesen, die für jeden Gattertyp die Umschaltcharakteristik einschliesslich der Leistungsaufnahme kennt. So kann eine sehr exakte Abschätzung der Leistungsaufnahme erfolgen, deren Grundlage das Anwenderprogramm ist.

Da der Leakagestrom auch von den Eingangswerten eines Gatters abhängig ist, wird dieser mit Hilfe von “PrimeTime” ebenso genauer bestimmt als im Synthesetool. Während für Strukturgrößen bis 250nm diese Unterschiede kaum zum Tragen kommen, sind diese bei 130nm und kleiner unter Umständen erheblich (vgl. Abb. 2).

5.2 Speicher

Microcontroller besitzen flüchtigen und nichtflüchtigen Speicher. Ist der flüchtige Speicher nicht mit Flipflops implementiert, wird mit Hilfe eines technologieabhängigen Generators jeweils ein Hardmacro[4] vom Speicher erzeugt. Dazu wird ein Verhaltensmodell, ein Layout sowie eine Netzliste für die Verifikation des Layouts generiert. Diese Netzliste, z.B. im SPICE Format, kann für eine analoge Simulation des Speichers benutzt werden. Auf diese Art kann sehr genau die Energie z.B. für einen Schreib- oder Lesezugriff ermittelt werden. Der zeitliche Aufwand dafür ist sehr hoch, so dass tatsächlich nur wenige Testfälle simuliert werden können und von diesen aus auf die durchschnittliche Leistungsaufnahme je Operation geschlossen werden. In Abb. 5 ist das Ergebnis einer beispielhaften Simulation eines SRAM Modells zu sehen.

Die nichtflüchtigen Speicher funktionieren nach dem Prinzip, dass ein physikalischer Effekt wie die permanente Widerstandsänderung ausgenutzt wird. Dieses ist nicht mit SPICE simulierbar. Anstelle dessen muss ein VerilogA Modell benutzt werden, welchse diese Aufgabe übernimmt. Dieses VerilogA Modell wird ausschließlich für die Simulation des veränderlichen Widerstands eingesetzt. Alle anderen Komponenten des Speichers werden, wie gehabt, als SPICE Netzliste simuliert. Somit ist auch eine genaue Simulation der Leistungsaufnahme dieses Speichertyps möglich.

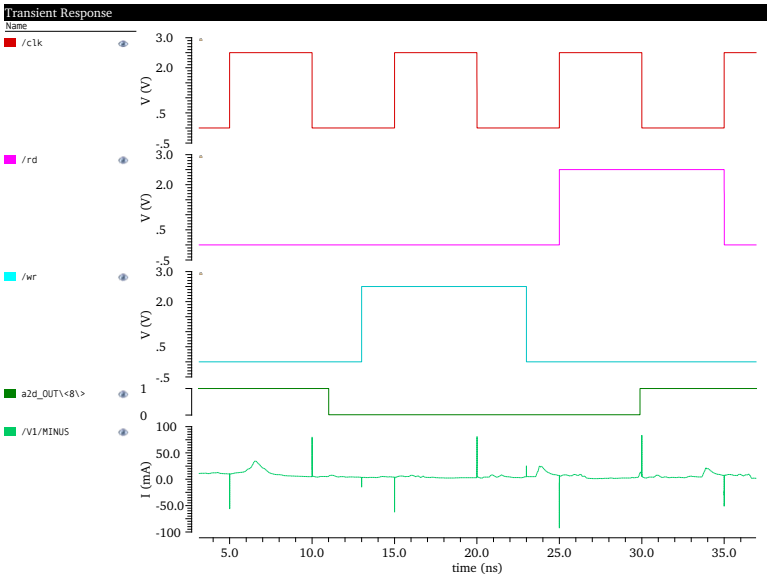


Abb. 5: Analoge Simulation eines SRAM

[4] Fertiges Modul auf Layoutebene

6 Simulationswerte

Da, wie in Abschnitt 5, beschrieben die Simulation der Leistungsaufnahme eine sehr gute Abschätzung des Energieverbrauchs liefert, kann auf dessen Grundlage der Speicher ermittelt werden, der am besten für den geplanten Anwendungsfall geeignet ist.

In Tab. 1 sind die Simulationswerte für die verschiedenen Speichertypen dargestellt. Für alle Speichertypen wurden mit auf Basis der SPICE Netzliste je ein Schreib- und Lesevorgang simuliert. Die Abb. 5 zeigt beispielhaft die grafische Darstellung des Simulationsergebnisses für einen SRAM. Die oberen vier Vektoren sind Clock, Lese- und Schreibsignal sowie der Ausgang eines ausgewählten Datenbits. Der untere Vektor ist die Leistungsaufnahme des Schaltkreises. Zur besseren Vergleichbarkeit sind alle Werte, unabhängig von der eigentlichen Speichergröße, für ein Bit angegeben. Für Flash und RRAM sind keine Werte für die Leckströme angegeben, da diese im Schlafmodus vollständig von der Spannungsversorgung getrennt werden können und im Betrieb die Leckströme eine untergeordnete Rolle spielen.

Speichertyp	Energie (Schreiben)	Energie (Lesen)	Leakage
Flipflop in Registerfile	0,195pWs	0,07pWs	250 pW
SRAM (aus 2kB)	10,7pWs	10,03pWs	45 pW
Flash (aus 64kb)	13400pWs	31,25pWs	0 pW
RRAM (aus 4kbit)	2pWs	31pWs	0 pW

Tab. 1: Energie zum Lesen/Speichern eines Bit für verschiedene Speichertypen in einer 0.25μm Technologie

Die Evaluierung aus Tab. 1 ergibt, dass in jedem Fall der Programmspeicher in Flash-Technologie durch einen RRAM ersetzt werden kann zugunsten der Energiebilanz. Während die Energie zum Lesen etwa gleich ist, benötigt der Flash ein vielfaches an Energie zum Abspeichern eines Bits. Dies ist durch die unterschiedlichen Ansätze bei der Speichertechnologie bedingt. Beim Flash ist der Transistor mit einem doppelten Gate ausgestattet und für die Platzierung der Elektronen zwischen den beiden Gates ist eine hohe Spannung für einen definiert langen Zeitraum notwendig. Beim RRAM dagegen wird ein Widerstand mit geringer Spannung und in vergleichsweise kurzer Zeit verändert.

Ein mögliches Szenario ist auch der Ersatz des SRAM, der als Datenspeicher dient. Während die Energie zum Schreiben eines Werte im SRAM höher ist als für den RRAM (vgl. Tabelle 1), ist die Energie zum Lesen deutlich geringer. Es gibt Anwendungen, wie Verschlüsselungsalgorithmen, welche eine Lookup-Tabelle benutzen. Diese ist im Datenspeicher abgelegt. Auf diese Daten wird stets lesend zugegriffen, so dass ein SRAM energiesparender sein kann als ein RRAM. Im Gegensatz zu Anwendungen, die weitgehend lesend auf den Datenspeicher zugreifen, gibt es Anwendungen die wechselweise lesend und schreibend den Datenspeicher nutzen, wie z.B. Sortieralgorithmen. In diesem Fall ist der Einsatz des RRAM sinnvoll. Neben der Betrachtung der Energie für die Lese- und Schreibzyklen ist auch der Leckstrom zu berücksichtigen. Unter Umständen ist der Energieaufwand für das Schreiben und Lesen des RRAM höher als beim SRAM. Dieses kann über die Zeit, die der Knoten im Schlafmodus ist, kompensiert werden. Anhand der Formel 1 kann der Anwender abschätzen, ob der Aufwand für den Einsatz eines RRAM gerecht-

fertigt ist, dabei ist der Speicher Typ 1 der SRAM und Typ 2 der RRAM. Mit der Annahme, dass einmal pro Aktivphase auf den Datenspeicher lesend und schreiben zugegriffen wird, ergibt sich eine Zeit von etwa 0,3s, die der Knoten in Ruhemodus sein muss, damit der höhere Energieaufwand für die Nutzung des RRAM kompensiert ist. Finden zum Beispiel in der Aktivphase ein Schreibzugriff und fünf Lesezugriffe statt, so ist eine Schlafzeit von etwa 2s notwendig. Ist die notwendige Dauer des Ruhemodus zur Kompensierung des höheren Energieaufwands größer als die Anwendung in ihrem Einsatzgebiet zulässt, sollte auf den RRAM verzichtet werden, da so die Energieeffizienz nicht verbessert werden kann.

Wenn der Knoten in den Schlafmodus übergeht, muss der Inhalt des Registerfiles im Schattenregister gespeichert werden. Das bedeutet das zunächst das Datum aus dem Registerfile gelesen (1) und ins Schattenregister geschrieben werden muss (2). Beim Aufwachen wird der Wert aus dem RRAM gelesen (3) und in das Registerfile zurückgeschrieben (4). Also müssen bei der Betrachtung der Energiebilanz alle vier Vorgänge berücksichtigt werden. Dagegen steht die Einsparung des Leckstroms. Die Formel 2 beschreibt diesen Vorgang, wobei Speicher Typ 1 das Flipflop-basierte Registerfile und Typ 2 das Schattenregister in RRAM Technologie ist. Mit dem Einsetzen der Parameter ergibt sich die Zeit für den Ruhemodus, ab welcher sich der Einsatz eines Schattenregister aus energetischer Sicht lohnt.

Mit den Werten aus der Simulation (Tab.1) benötigen alle Vorgänge zusammen ca. 33,3 pWs um ein Bit zu speichern und wieder abzurufen. Unter Berücksichtigung des Leckstroms eines Flipflops von 250pW, lohnt sich es ab einer Schlafdauer von 0,13s auf den nichtflüchtigen Speicher zurückzugreifen. Nimmt man den Energieaufwand für einen Controller dazu, welcher das Management des Datentransfers übernimmt, kann man annehmen, dass sich ab einer Schlafdauer im Sekundenbereich der Einsatz von RRAM als Zwischenspeicher sinnvoll erscheint.

$$T = \frac{(n \times E_{Lesen_{Typ_1}} + m \times E_{Schreiben_{Typ_1}}) - (n \times E_{Lesen_{Typ_2}} + m \times E_{Schreiben_{Typ_2}})}{P_{Leckstrom_{Typ_1}}} \tag{1}$$

$$T = \frac{E_{Lesen_{Typ_1}} + E_{Schreiben_{Typ_2}} + E_{Lesen_{Typ_2}} + E_{Schreiben_{Typ_1}}}{P_{Leckstrom_{Typ_1}}} \tag{2}$$

wobei:

n = Anzahl der Lesezyklen in der Aktivphase
m = Anzahl der Schreibzyklen in der Aktivphase
$E_{Lesen_{Typ_1}}$ = Energie zum Lesen eines Bits von Speichertyp 1
$E_{Schreiben_{Typ_1}}$ = Energie zum Schreiben eines Bits von Speichertyp 1
$E_{Lesen_{Typ_2}}$ = Energie zum Lesen eines Bits von Speichertyp 2
$E_{Schreiben_{Typ_2}}$ = Energie zum Schreiben eines Bits von Speichertyp 2
$P_{Leckstrom_{Typ_1}}$ = Leckstrom eines Bit von Speichertyp 1

7 Zusammenfassung

In dieser Arbeit wurde die Möglichkeit betrachtet, neue nichtflüchtige Speicher in Microcontrollern vor dem Hintergrund von verlängerter Batterielebensdauer einzusetzen. Gerade in Anwendungsgebieten mit langen Standby-Zeiten kann so die Energiebilanz verbessert werden. Ab einer Schlafdauer von etwa einer Sekunde ist der zusätzliche Energieaufwand für das Puffern in einem nichtflüchtigen Speichern kompensiert. Wichtig dafür ist eine Applikation, bei der sich die Wach- und Schlafphasen gut abschätzen lassen. Somit werden negative Auswirkungen des zusätzlichen Speichers in RRAM Technologie auf den Energieverbrauch vermieden. Darüber hinaus ist es notwendig mit Hilfe von geeigneten Modellen und den zugehörigen Simulationswerkzeugen den Energieverbrauch für Schreib- und Leseoperationen sowie des Leckstroms der verschiedenen Speichertypen bestimmen zu können.

Literaturverzeichnis

[Gi09] Girard, Olivier: , openMSP430 microcontroller - compatible with the original MSP430 architecture. `http://opencores.org/project,openmsp430`, 2009. [Online; accessed 2015-03-05].

[IH15] IHP Microelectronics: , Desgin Kit. `http://www.ihp-microelectronics.com/en/services/mpw-prototyping/design-kit/design-kit.html`, 2015. [Online; abgefragt 24.4.2015].

[Le08] Lee, H.Y. and Chen, P.S. and Wu, T.Y. and Chen, Y.S. and Wang, C.C. and Tzeng, P.J. and Lin, C.H. and Chen, F. and Lien, C.H. and Tsai, M.J.: Low Power and High Speed Bipolar Switching with a Thin Reactive Ti Buffer Layer in Robust HfO2 Based RRAM. In: Electron Devices Meeting, 2008. IEDM 2008. IEEE International. S. 1–4, Dec 2008.

[NB64] Nielsen, P.H.; Bashara, N.M.: The Reversible Voltage-Induced Initial Resistance in the Negative Resistance Sandwich Structure. Electron Devices, IEEE Transactions on, 11(5):243–244, May 1964.

[Pa14] Panic, Goran: A Methodology for Designing Low Power Sensor Node Hardware Systems. Dissertation, Brandenburgische Technische Universität Cottbus-Senftenberg, 2014.

[Pa15] Panasonic: , MN101L series microcontroller. `http://www.semicon.panasonic.co.jp/en/products/microcomputers/mn101l`, 2015. [Online; abgefragt 27.4.2015].

[Sy15] Synopsys: , PrimeTime. `http://synopsys.com/Tools/Implementation/SignOff/PrimeTime/Pages/default.aspx`, 2015. [Online; abgefragt 27.4.2015].

[Te15a] Texas Instruments: , Low-power MCUs. `http://www.ti.com/msp430`, 2015. [Online; abgefragt 27.4.2015].

[Te15b] Texas Instruments: , MSP430x5xx and MSP430x6xx Family User's Guide. `http://www.ti.com/lit/ug/slau208n/slau208n.pdf`, 2015. [Online; abgefragt 27.4.2015].

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Workshop Industry 4.0: Informatics enables new Production Systems

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Workshop Industry 4.0: Informatics enables new Production Systems

Rolf Kraemer[1],Thomas Haustein[2]

Abstract: The workshop addresses the important area of production and manufacturing technologies, which faces today another industrial revolution called „Industry 4.0“. The strong connectivity of systems causes on one hand side completely new chances for intelligent fabrication control and for the integration into logistic processes, and on the other hand side new challenges to the IT systems and risk management. The workshop addresses these topics in a vertical approach such that different topic from different technology areas can be put into perspectives.The workshop consists of 5 invited presentations that give overviews of the addressed topic and 6 reviewed submissions

Programm Committee:

Prof. Rolf Kraemer (IHP, BTU Cottbus-Senftenberg)

Dr. Thomas Haustein (Fraunhofer-HHI Berlin)

Prof. Peter Langendörfer (IHP, BTU Cottbus-Senftenberg)

Prof. Eckhard Grass (IHP, Humboldt-Universität zu Berlin)

Prof. Björn Scheuermann (Humboldt-Universität zu Berlin)

Prof. Hans Schotten (DFKI, University Kaiserslautern)

Dr. Andreas Vedral (WAGO)

Dr. Karsten Walther (Harting)

Prof. Rüdiger Kays (TU Dortmund)

Prof. Oliver Wetter (FH Bielefeld)

Prof. Jörg Wollert (FH Aachen)

1 Innovations for High Performance Microelectronics in Frankfurt (Oder) and BTU - Brandenburg University of Technology Cottbus-Senftenberg, Drahtlose Systeme, Im Technologiepark 25, 15236 Frankfurt-Oder, kraemer@ihp-microelectronics.com

2 Fraunhofer Heinrich Hertz Institute, Drahtlose Kommunikation und Netzwerke,,Einsteinufer 37,10587-Berlin, thomas.haustein@hhi.fraunhofer.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Big Data, Industrie 4.0 und Co.: Wie sich die Welt verändert.

Peter Liggesmeyer[1],

Abstract: Industrie 4.0 bezeichnet die vierte industrielle Revolution. Die erste industrielle Revolution war charakterisiert durch die Einführung der Dampfmaschine als neue Querschnittstechnologie. Diese Entwicklung hat fast alle Lebensbereiche drastisch verändert. Die zweite industrielle Revolution ist verbunden mit dem Namen Henry Ford. Mit ihr wurde die Massenproduktion durch die Einführung von Takt und Band ermöglicht. Fertigungsprozesse wurden in viele kleine Schritte zerlegt, einhergehend mit einer starken Standardisierung der Produkte. Die Massenproduktion von Gütern führte zu einer Preisreduktion, allerdings auf Kosten der Individualität, da Produkte in immer gleichen Prozessen gefertigt wurden. Die dritte industrielle Revolution war in den 70er und 80er Jahren des letzten Jahrhunderts die weitgehende Automatisierung der Produktion. Manuelle Fertigungsschritte wurden zunehmend von computergesteuerten Maschinen übernommen, die Produkte schneller, präziser und in höherer Qualität fertigten, aber eben nach wie vor vereinheitlichte Massenprodukte. In der vierten industriellen Revolution geht man nun davon aus, dass man das Massenprodukt durch massenindividualisierte Produkte ersetzt, die an den Bedürfnissen des Kunden ausgerichtet sind. Die Themen Industrie 4.0 und Big Data sind eng miteinander verwoben. Technisch gesehen kann man Big Data als einen wichtigen Aspekt der informationstechnische Seite zu Industrie 4.0 verstehen. Die Daten müssen in die Produktionsumgebung hineinkommen und dort korrekt verarbeitet werden. Der Produktionsprozess läuft nicht mehr schrittweise in der immer gleichen Weise ab, denn dann käme ja immer das gleiche Produkt dabei heraus. Die Produktionsumgebung muss vielmehr die Fähigkeit haben, auf sich ändernde Anforderungen einzustellen. Das kann nicht mehr wie bisher durch einen Menschen durchgeführt werden, sondern muss autonom geschehen. Die massenindividualisierte Produktion benötigt einen hohen Grad an Autonomie.

1 Fraunhofer Institute IESE, Fraunhofer-Platz 1, 67663 Kaiserslautern, peter.liggesmeyer@iese.fraunhofer.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Herausforderung:Vertikale Integration – Vom Shop Floor zum Top Floor

Claus Hilger[1],

Abstract: Die Vision der Industrie 4.0 umfasst einerseits cyberphysische Systeme im Feld und andererseits zentrale leistungsfähige ERP Systeme. Technisch scheint der Weg dahin einigermaßen klar zu sein. Man benötigt objektidentifizierende Sensorik und Aktuatorik im Feld, um einen konkreten technischen Vorgang einem übergeordneten Prozess zuordnen zu können. Dazu kann z.B. RFID genutzt werden. Weiterhin werden Kleinst-Rechner im Feld benötigt, welche adaptiv die Vorgaben des übergeordneten Prozesses umsetzen aber auch selbst Entscheidungen treffen können. Diese „intelligenten" Geräte übernehmen auch eine Vorverarbeitungsfunktion von Daten, welche in das zentrale System hochgeladen werden. Im Top Floor werden Warenwirtschafts Produktionsprozesse feingranular verfolgt und gesteuert aber auch Daten gesammelt und anlagenübergreifend mit Blick auf vorauseilende schauende Wartung ausgewertet.
Die große Herausforderung dabei ist, die dynamische Verteilung von Entscheidungen und Kommunikation über mehrere Ebenen und die Kommunikation heterogener, dynamisch miteinander verbundener Systeme. Hier erscheinen Middleware Architekturen unerlässlich. Zudem ist es insgesamt schwer vorstellbar, lauffähige Gesamtsysteme ohne tiefes Breitenwissen umzusetzen. Dies ist nicht nur eine Herausforderung für den Werker Mitarbeitenden an der Maschine, der das Verhalten der Maschine zu jeder Zeit nachverfolgen können muss, sondern auch für die Systemintegration, welche diese Systeme konzipiert. Nur interdisziplinäre Teams, welche alle Ebenen abdecken, werden dieser Aufgabe gewachsen sein.

[1] HARTING IT Services GmbH & Co. KG, Abteilung, Marienwerderstraße 2, 32339 Espelkamp, claus.hilger@HARTING.com

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Smart Factory based on intelligent technical systems

Jürgen Jasperneite[1]

Abstract: Customers increasingly want to have individual products with costs comparable to mass-produced goods. For manufacturers this leads to small batches and many variants associated and a high productivity at the same time. Therefore future production systems must be adaptive, resource-efficient and user friendly. Such a smart factory is an intelligent socio-technical production system based on informed people, informed machines and informed products. For the smart factory the usage of Information and Communication technologies (ICT) and intelligent automation is of high importance. Especially a seamless connectivity, the efficient computer modeling of knowledge, knowledge-based algorithms for self-x capabilities and user-friendly and intuitive interaction technologies are ingredients to realize the smart factory paradigm. This talk shows how the needed intelligence comes into the technical systems at the shop floor level. Some examples of the leading edge technology cluster "intelligent technical systems OstWestfalen-Lippe it's OWL" are introduced. In this cluster, which is an integral part of germans high-tech strategy Industrie 4.0, 178 partners from research and industry are working on solutions for the smart factory and smart products. Based on real-world industrial applications methods and technologies for the self-configuration and self-optimization of machines and systems as well as the usage of advanced human-computer interaction technologies are presented.

[1] Fraunhofer IOSB-INA and Institute for Information Technologies of OWL University of Applied Sciences, Lemgo, juergen.jasperneite@iosb-ina.fraunhofer.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Fast Actuators Sensors and Transceivers (FAST)

Frank Ellinger[1]

Abstract: We will give an overview of the German zwanzig20 BMBF cluster FAST which aims at achieving a technological and economic breakthrough by means of real-time capabilities. The FAST consortium consist of 44 small and medium sized enterprises, 19 large companies, 8 universities, 6 research institutes and one association. The FAST partners encompass the whole competency and value chain from materials, semiconductors, components and software to complex systems and communication networks. FAST comprises around 20 projects and is scheduled from 2014 to 2020.
For many applications in the area of security, traffic, medical, sports, consumer electronics and communications, real-time capabilities enable new application functionality and markets. The entire process including sensing, data transfer, processing and actuation must have a negligible delay. This is in particular a huge challenge for systems that connect a high amount of sensors and actuators comprehensively via mobile communication networks. In this regard, FAST wants to get as close as possible to the ultimate physical limitation given by the speed of light. In ideal case, a signal needs only 0.1 ms for 30 km.
Examples for specific application of the real-time technologies are: 5G communications, automated driving, road traffic management, 10 Gb/s real-time Ethernet in vehicles, real-time radar, sensor networks, real-time automated industry, real-time control, testing and synchronization of machines, motors and work pieces, movement analysis and optimization of athletes and in the care sector, tele-surgery, remote treatment of patients, virtual reality apps, real-time connection of musicians via internet, and real-time cloud services.
The authors acknowledge the BMBF for financial support of FAST.
Moreover, we will present research of our chair in the area of high speed and low power consuming systems, transceivers and circuits.

[1] Technische Universität Dresden, Schaltungstechnik und Netzwerktheorie, 01062 Dresden, Frank.Ellinger@tu-dresden.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

5G und Industrie 4.0

Hans D. Schotten[1]

Abstract: Mit der zunehmenden Akzeptanz von Industrie 4.0 als Leitvision für die Weiterentwicklung von Produktion und Logistik werden zurzeit auch viele andere Wirtschaftsbereiche identifiziert, in denen deutliche Effizienzsteigerungen durch Digitalisierung und Vernetzung erwartet werden können. In diesem Zusammenhang kann der 5. Generation mobiler Kommunikationsnetze (5G) eine besondere Bedeutung zukommen, da diese mit dem erklärten Ziel designt wird, „professionelle" Dienste zu unterstützen. Um die hier erhofften Synergien aber tatsächlich realisieren zu können, müssen Funktionalitäten, Architekturen und Roadmaps aufeinander abgestimmt werden.
Dieser Beitrag untersucht Herausforderungen und Chancen, die sich aus der parallelen und nur vorsichtig verzahnten Entwicklung der beiden technischen Leitinnovationen 5G und Industrie 4.0 bzw. Digitalisierung der Wirtschaft ergeben. Er versucht, Beiträge zur Identifikation von Abstimmungsbedarf zu leisten. Neben den technischen Herausforderungen werden hierbei auch Fragen der Regulierung, Netzneutralität und Kompatibilität von Geschäftsmodellen angesprochen.

[1] DFKI GmbH und Technische Universität Kaiserslautern, 67663 Kaiserslautern, schotten@eit.uni-kl.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Einfluss von Industrie 4.0 auf Wertschöpfungsnetze am Beispiel der Textilreinigungsbranche

Ulrich Berger[1,2], Kornelius Wächter[1,3] und Janny Klabuhn[1,4]

Abstract: In dem folgenden Artikel wird der Einfluss des Industrie 4.0-Ansatzes [Zu14] auf die Wertschöpfungsnetze der Textilindustrie analysiert. Es werden Effizienzsteigerungen und weitere Chancen der durch den Einsatz der in der Industrie 4.0 verfügbaren Technologien erörtert. Als Beispiel fungiert die Textilreinigungsbranche. Durch den Einsatz von RFID-Technologien soll es zu einer signifikanten Erhöhung der Effizienz und Produktivität kommen [En15]. Es wird die RFID-Technologie eingesetzt, um jedes einzelne Textil zu identifizieren, wodurch es zu einer stück- und artikelgenauen Erfassung aller Wäscheteile kommt. Als zusätzlicher Nutzen für die Wäscherei und den Kunden dient die einfache und schnelle Mengenbestimmung des Wareneingangs durch Pulk-Erfassung. Außerdem führen die Auftragsverfolgung und die Mengeninformationen zu einer Optimierung der Prozesse in der Wäscherei.

Keywords: Industrie 4.0, RFID, Supply Chain Textilindustrie, Chancen, Textilreinigungsbranche

1 Supply-Chain Textilreinigungsbranche

In der Textilreinigungsbranche ist es sehr verbreitet Hotels und anderen Großkunden, Wäsche zur Miete anzubieten. Dabei besitzt die Wäscherei ein bestimmtes Spektrum an Textilien, wie z. B. Bettwäsche oder Handtücher und vermietet diese an die Kunden. Der Kunde bezieht die Dienstleistung, saubere Wäsche angeliefert zu bekommen und nach Nutzung wieder abgeben zu können. In der Regel haben Hotels einen enorm hohen Wäschedurchsatz, weshalb es ihnen meist nicht möglich ist, die Textilien intern zu reinigen. Das Mietwäschemodell bietet den Kunden den großen Vorteil, dass sie nur kleine Lager mit Wäsche betreiben müssen und je nach Auftragslage unterschiedliche Mengen an Frischwäsche beziehen können. Außerdem kommt es zu einer Reduktion von gebundenem Kapital.

In Abbildung Nr. 1 wurde die Supply-Chain des Mietwäschemodells aus der Perspektive einer Wäscherei dargestellt. Die Güter verlaufen innerhalb der abgebildeten Wertschöpfungskette von links nach rechts. Als Güter werden dabei die unterschiedlichen Textilien

[1] Brandenburgische Technische Universität Cottbus-Senftenberg, Lehrstuhl Automatisierungstechnik, Siemens-Halske-Ring 14, 03046 Cottbus
[2] berger.u@b-tu.de
[3] kornelius.waechter@b-tu.de
[4] janny.klabuhn@b-tu.de

verstanden, welche sich in der Regel in Trocken- und Mangelwäsche unterscheiden lassen. Zu der Trockenwäsche gehören alle Textilien, denen in großen Lufttrocknern das Restwasser des Waschvorgangs entzogen wird. Dazu gehört exemplarisch Frotteewäsche, wie z.B. Handtücher. Mangelwäsche unterteilt sich in Bettwäsche, Tischwäsche und Kleinteilwäsche.

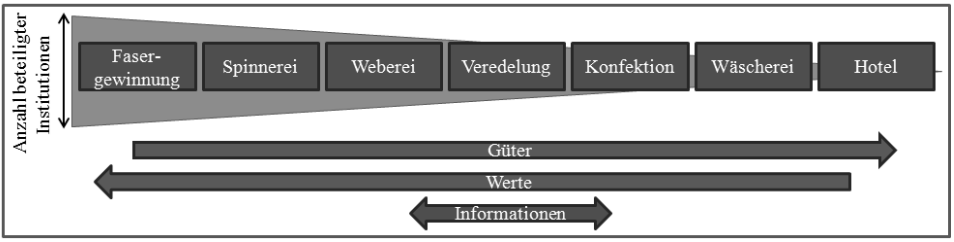


Abb. 1: Supply-Chain Großwäscherei

Die gereinigten Güter werden je nach Auftragslage durch den Kunden, in diesem Beispiel die Hotelbranche, in unterschiedlichen Mengen bestellt. Die Wäschereien müssen über genügend große Lager verfügen, um mehreren Kunden flexibel, ausreichend Textilien in guter Qualität anbieten zu können. Aufgrund der hohen Belastung durch den regelmäßigen Reinigungsprozess, müssen die Produkte nach sehr kurzer Lebensdauer ersetzt werden. Zudem kommt es zu einem sehr hohen Schwund von Textilien innerhalb der Hotelbranche, welcher kompensiert werden muss. Deshalb werden regelmäßig neue Textilien bei Großhändlern oder den Herstellern direkt bezogen. Diese Textilhersteller haben sich meistens nur auf das Design, die Konfektion und den Vertrieb spezialisiert und arbeiten oftmals mit Subunternehmern zusammen. Die Stoffe für die Textilien werden durch andere Firmen nach den Wünschen der Textilhersteller veredelt, was bedeutet, dass diese chemisch oder organisch gefärbt oder bearbeitet werden. Die Stoffe wiederum werden von Webereien aus verschiedenen Garnen hergestellt. Diese werden zuvor aus chemischen oder organischen Fasern in Spinnereien gefertigt. Am Anfang der Wertschöpfungskette steht die Fasergewinnung, welche auf Kunstfasern, Baumwolle, Hanf oder Wolle beruht. Diese Wertschöpfungskette ist in die Richtung der Produzenten von breiterer Struktur, wodurch eher von einem Wertschöpfungsnetz, mit sehr komplexen Vernetzungen und Abhängigkeiten, zu sprechen ist.

Das Kapital, fließt entgegen der Güter in Richtung der Produzenten. In der Textilindustrie gibt es kaum Firmen, die sich entlang der gesamten Wertschöpfungskette ausgebreitet haben, da die einzelnen Segmente sehr spezialisiert sind. Daher gibt es viele einzelne Firmen, die einen Teil des Kapitalflusses abzweigen. Oftmals stehen zwischen den Segmenten Zwischenhändler. Diese verfolgen ebenfalls wirtschaftliche Interessen wie beispielsweise Bestandssicherung, Weiterentwicklung, Kundengewinnung und Erlangen von Wettbewerbsvorteilen am Markt.

Der Informationsfluss bezüglich der Bedarfe der Endverbraucher ist aufgrund der komplexen Strukturen nur sehr schwach ausgeprägt und nimmt in Richtung der Produzenten ab. Meistens gelangt der Großteil der Informationen nicht weiter als von einem Segment zu dem nächsten. Daher ist es für den Kunden sehr schwierig zu erfahren, woher das Produkt stammt bzw. welche Arbeitsbedingungen während der Produktion herrschten und ob umweltschonende Produktionsverfahren eingesetzt wurden.

Dennoch gelangen immer noch mehr Informationen vom Hersteller zum Konsumenten als in die andere Richtung. Folge der noch weniger ausgeprägten Kommunikation in diese Richtung ist der Peitscheneffekt. Dabei werden durch die mangelnde Informationsübertragung vom Kunden zum Anbieter, die Lager- und Produktionsschwankungen immer größer, weshalb es zu einem Aufschaukeln der Lagermengen entlang der gesamten Lieferkette kommt. Dieser Vorgang wird in der Abbildung 2 verdeutlicht.

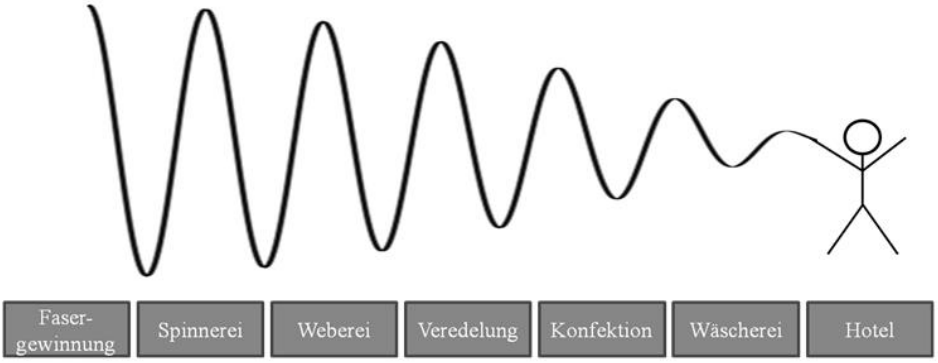


Abb. 2: Peitscheneffekt Textilbranche

Die Ursache für den Peitscheneffekt ist, dass die beobachtete Nachfrage als zukünftige Nachfrage aufgefasst wird. Bei Bestellungen kommt es oftmals zu einer Auftragsbündelung, um Mengenrabatte zu erlangen oder um einen optimalen Einstandskurs bei Gütern mit volatilen Preisen zu erlangen. Außerdem werden Aufträge oftmals zur Reduktion von Lieferkosten gebündelt.

In der Hotelbranche gibt es häufig ein Wäschelager auf jeder Etage sowie ein Hauptwäschelager. Da die Zimmer nicht immer gleichmäßig ausgebucht sind und es zu Schwankungen des Wäscheverbrauchs je nach Saison und Kunde kommt, werden unterschiedlich viele Textilien aus den Wäschelagern auf den Etagen durch die Reinigungskräfte entnommen. Wenn diese Wäschelager aufgefüllt werden, kommt es zu einer hohen Nachfrage an Textilien im Hauptwäschelager. Damit die Nachfrage mehrerer Etagenlager zu jeder Zeit bedient werden kann, muss das Hauptwäschelager über dementsprechend viele Textilien verfügen. Dieses wird wiederum von der Wäscherei mit frischer Wäsche versorgt. Die Wäscherei geht bei einer hohen Nachfrage an Textilien davon aus, dass zukünftig ebenfalls eine hohe Nachfrage bestehen wird und passt die Lagergröße dementsprechend an. Dieser Effekt tritt ebenfalls auf, wenn neue Wäsche von den Her-

stellern bezogen wird, um Schwund und Abnutzung auszugleichen bzw. um die Lagermenge anzupassen. Je weiter sich dieser Prozess entlang der Lieferkette zieht, desto extremer schaukelt sich die Nachfrageschwankung auf. Der Grund dafür ist, dass der Kunde immer eine sofortige Befriedigung seiner Nachfrage beim Lieferanten erwartet und der Lieferant seine Lagergröße dementsprechend anpassen muss. Falls die Nachfrage nicht durch den Lieferanten abgedeckt werden kann, hat der Kunde durch die Wettbewerbssituation jederzeit die Möglichkeit den Lieferanten zu wechseln.

Die einzige Möglichkeit, um dem Peitscheneffekt zu entgehen, ist ein durchgängiger Informationsaustausch entlang der Wertschöpfungskette. Dabei müssen die Informationen über die Absatzmengen der Kunden zurück zum Lieferanten gelangen, damit dieser seine Lagerumfänge danach ausrichten kann. Damit Lagerdurchsätze und Lagermengen effizient aufgenommen werden können, muss jedes Produkt über eine eigene Identität verfügen und Informationen speichern können. Durch eine komplette Vernetzung der Segmente einer Wertschöpfungskette sowie die Informationsübertragung durch das Produkt und das Internet der Dinge und Dienste, kann es zu enormen Effizienzsteigerungen kommen.

2 Chancen des Einsatzes von Industrie 4.0 in der Textilindustrie

Durch den Einsatz von Industrie 4.0 in der Textilreinigungsbranche ergeben sich zahlreiche Vorteile. Ein wesentlicher Punkt ist der verbesserte Austausch von Informationen entlang der Supply-Chain. Aufgrund der Tatsache, dass die Segmente der Wertschöpfungskette jederzeit über die Lagermengen der Kooperationspartner, durch das Betreiben eines einheitlichen segmentübergreifenden Informationssystems, verfügen, können diese wirtschaftlicher zusammenarbeiten. Dabei wird der Peitscheneffekt deutlich geschwächt und das an Güter gebundene Kapital reduziert. Außerdem wird es aufgrund der beschriebenen Szenarien durch den Einsatz der RFID-Technologie und der Automatisierung der Prozesse zu einer deutlichen Verringerung von Bearbeitungszeiten und zu einer klaren Effizienzsteigerung kommen. Die eindeutige Identifizierung jedes Textils ermöglicht die permanente Lokalisierung und verringert Schwund durch fehlerhafte Logistikabläufe.

Durch den Einsatz von RFID-Transpondern in der Textilindustrie kann jedes Produkt individuell nachverfolgt werden, wodurch mehr Transparenz für den Kunden entsteht. In diesem Fall soll der Transponder bereits während der Konfektion mit dem Textil verknüpft und mit den relevanten Daten beschrieben werden. Dazu zählen unter anderem Produktionsstandort, Bearbeiter, Ausgangsmaterial und dessen Ursprung, verwendete Hilfsmittel und Chemikalien. Die Produktinformationen werden an jedem Segment der Wertschöpfungskette ergänzt und durch das Textilunternehmen verifiziert. Es könnten sich Textilanbieter auf den Märkten etablieren, die diese absolute Transparenz für den Kunden, als Marketinginstrument nutzen.

3 Umsetzung Industrie 4.0 in Großwäschereien

Konkrete Umsetzungsideen für den Einsatz der Technologie im Umfeld von Großwäschereien werden derzeit im Rahmen eines durch das Bundesministerium für Wirtschaft und Technologie gefördertes Projekt mit zwei Wirtschaftspartnern unter Trägerschaft der VDI/VDE Innovation + Technik GmbH an der Brandenburgischen Technischen Universität Cottbus-Senftenberg erforscht.[2]

Um die Vorteile der RFID-Technologien innerhalb einer Großwäscherei nutzen zu können, werden die Textilien im Rahmen der Industrie 4.0 mit RFID-Transpondern ausgestattet. Diese dienen zur Speicherung von Informationen und zur einfachen Kommunikation mit den Textilien. Dabei werden die Transponder während der Konfektion oder innerhalb der Großwäscherei mit den Textilien verbunden. Die speziellen RFID-Tags für den Einsatz innerhalb von Wäschereien werden in der Regel an das Textil genäht, geklebt oder in Form eines Etiketts angebracht. Aufgrund ihrer guten Eignung für diesen Anwendungsbereich sind UHF-Transponder sehr verbreitet, andere RFID Systeme werden zum Einsatz gebracht, wenn spezielle Anforderungen vorliegen. Diese passiven UHF-Transponder können die verschiedensten Informationen auf dem Produkt speichern. Ein wesentlicher Punkt dabei, ist die sichere Identifikation des Objekts. Dazu gehören:

- die Artikelnummer,
- Typ des Objekts,
- eindeutige Produktnummer,
- Hersteller,
- Besitzer,
- Pflegehinweise und
- absolvierte und maximale Anzahl Waschvorgänge.

Im Folgenden findet sich eine Auflistung der wichtigsten Eigenschaften der in der Textilreinigungsbranche verbreiteten UHF-Transponder:

- Passive Transponder
- UHF-Übertragungsfrequenz (865-950 MHz)
- Pulk-Erfassung
- Lesedistanz bis zu sechs Metern
- Druckbeständigkeit in der Entwässerungspresse
- Temperatur- und Druckbeständigkeit beim Mangelprozess
- Einsatz für mindestens 200 Waschperioden
- Einstandspreis für Wäschereien und Großkunden unter 0,50 € pro Stück

[2] „Laundry Robotics Co-Techno/ RFID-System zur räumlichen Lokalisierung und Orientierungserkennung der Stellplätze von Wäschestapeln und Mitwirkung bei der Entwicklung eines sich der Stapelgröße anpassenden Greiferkorbs“ Kennzeichen: 16KN040622

Zum derzeitigen Stand der Technik existieren auf dem Markt verschiedene Anbieter von RFID-Technologien mit unterschiedlichen Spezifikationsstrategien. Die Herausforderung besteht darin, die für die zu untersuchenden Einsatzszenarien geeigneten Komponenten der bereits auf Textilien spezialisierten Anbieter zu identifizieren und mit weiteren Technologien zur Lösung der Herausforderungen in Großwäschereien zu verbinden.

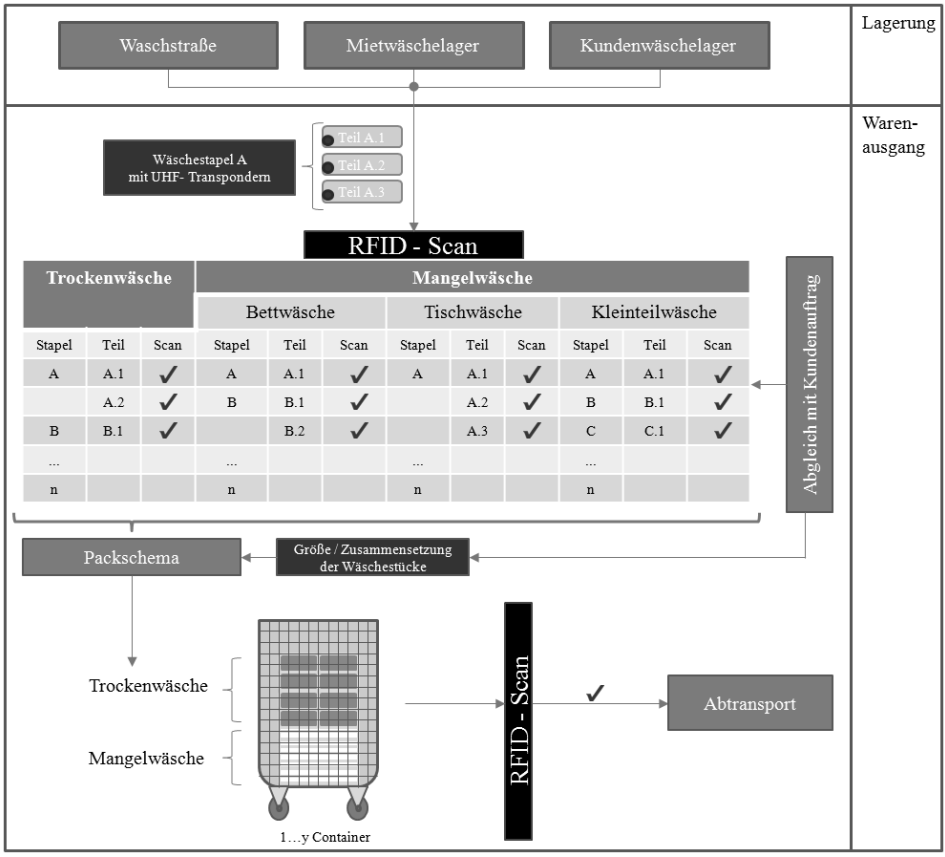


Abb. 3: Kommissionierung Wäscherei – Ablauf

Der Ablauf der Kommissionierung der Textilien innerhalb der Wäschereien ist in Abbildung 3 visualisiert. Die sauberen Gewebe werden nach dem Reinigungsprozess zunächst eingelagert. Die Textilien werden entweder in dem Lager der Waschstraße, dem Mietwäschelager oder dem Kundenwäschelager platziert. In den jeweiligen Lagern befinden sich Regale mit den gestapelten Wäschestücken, wobei jedes Stück mit einem eigenen UHF-Transponder ausgestattet ist. Wenn ein Kunde eine bestimmte Menge an Wäsche bestellt, werden die benötigten Stapel automatisiert aus den Lagern entnommen und auf Grundlage des Kundenauftrags zusammengestellt. Dabei wird per RFID überprüft, ob

alle Stapel und dazugehörigen Textilien vorhanden sind. Auf Basis der Zusammensetzung des Kundenauftrags wird per Computersteuerung ein optimales Packschema für die Wäschecontainer ermittelt, wobei die Mangelwäsche mit der höheren Dichte aus Stabilitätsgründen unter der Trockenwäsche platziert werden muss. Bevor die bestückten Ladungsträger die Wäscherei verlassen, werden sie durch ein mit RFID-Lesegeräten bestücktes Tor geschoben. Dabei wird per Pulk-Erfassung überprüft, ob alle Komponenten des Kundenauftrags vorhanden sind.

In der Abbildung 4 wurde ein exemplarischer Aufbau der Kommissionierung in der Draufsicht dargestellt. In diesem Fall werden die Wäschestapel aus einem der Lager per Transportband angeliefert. Ein Roboter übernimmt auf Grundlage der aus den RFID-Daten zur Verfügung gestellten Geometrieeigenschaften des Wäschestapels das Einsortieren in den Container. Ein führerloses Transportsystem befördert diese anschließend durch ein RFID-Tor zum Versand.

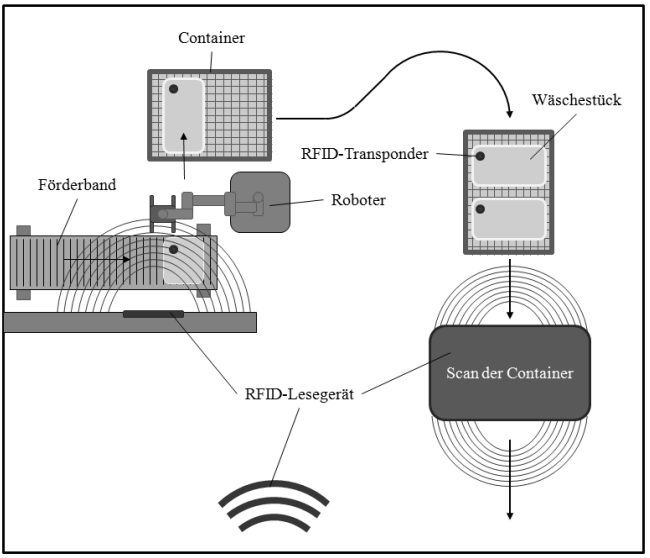


Abb. 4: Kommissionierung Wäscherei – Aufbau

Sobald die saubere Wäsche bei den Kunden eintrifft, wird die Lieferung durch ein mobiles RFID-Lesegerät auf Vollständigkeit überprüft. In der Hotelbranche kommt es anschließend zu einem Einsortieren in das Hauptlager. Sobald Wäsche aus dem Hauptlager oder den Wäschelagern auf den Etagen entfernt oder einsortiert wird kommt dieses mobile RFID-Lesegerät zum Einsatz. Dadurch ist es jederzeit möglich die exakten Lagermengen zu bestimmen bzw. zu aktualisieren.

Die Schmutzwäsche wird in regelmäßigen Abständen zurück zur Wäscherei befördert und per Pulk-Erfassung eingelesen. Anschließend kommt es zu einer automatischen Trennung der Textilien nach Waschverfahren, auf Grundlage der Informationen auf dem

Transponder. Die Bearbeitungsmaschinen können dabei individuell auf die Bearbeitungshinweise eingehen und es kommt zu einer direkten Kommunikation zwischen den Textil-Tags und den Maschinen.

4 Fazit

Die vierte industrielle Revolution bietet enorme Potentiale, um die Segmente der Wertschöpfungsketten miteinander zu vernetzen und dadurch den Informationsaustausch zu beflügeln. Demgemäß kann es zu signifikanten Effektivitätssteigerungen entlang der Supply-Chain kommen. Durch die RFID-Systeme und deren Einfluss auf die Umgebung, kommt es zu einer kompletten Änderung der Planungen innerhalb der Industrie. Die Ausrichtung der Produkte kann durch die Industrie 4.0 wesentlich stärker auf die Bedarfe der Kunden angepasst werden, weshalb die gesamte Produktion flexibel aufgestellt sein muss.

Für die Textilindustrie bietet der Einsatz von RFID-Technologie den großen Vorteil, dass der Kunde, durch die gespeicherten Informationen auf den Textilien, die Möglichkeit erhält, seinen Konsum auf der Grundlage von Informationen zu den Herstellungsbedingungen zu gestalten. Neben den sozialen und ökologischen Aspekten, birgt Industrie 4.0, durch die vollkommene Vernetzung entlang der gesamten Supply-Chain, enorme ökonomische Potenziale.

Handlungsbedarfe werden im Zusammenhang mit den beschriebenen Herausforderungen und Forschungen darin gesehen, die verschiedenen verfügbaren Informationssysteme und Technologien über geeignete kooperierende Schnittstellen nutzbar zu machen. Eine derartige Ausrichtung bietet Raum für zukünftige Innovationen und Potenzial für eine verbesserte Implementierung von Industrie 4.0 in den Prozessen von mittelständischen Unternehmen [Um13].

Literaturverzeichnis

[En15] Entwicklung eines Verfahrens und der damit verbundenen logistischen Abläufe zum ganzheitlichen RFID-Einsatz im geschlossenen Wäscherei-Kreislauf unter Einbeziehung der Prozessabläufe bei Großkunden, http://w- lin.eu/gruen/abschlussbericht_rfid_online.pdf, 13.05.2015.

[Um13] Umsetzungsempfehlung für das Zukunftsprojekt Industrie 4.0. Promotorengruppe Kommunikation der Forschungsunion Wirtschaft - Wissenschaft, 2013.

[Zu14] Zukunftsbild „Industrie 4.0“. Bundesministerium für Bildung und Forschung, 2014, Bonn.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Generische Funksysteme für IoT und Industrie 4.0

Jörg F. Wollert[1],

Abstract: Eingebettete Systeme sind aus keinem Lebensbereich mehr Weg zu denken. Von der Kaffeemaschine über Smart-Watches, bis zu technischen Systemen die uns das Leben vereinfachen. Embedded Systems sind allgegenwärtig. Die Informationen, die gesammelt und verarbeitet werden, sind vielfach so interessant, dass sie ausgetauscht und aufbereitet werden. Am besten frei von allen Kabeln und technischen Einschränkungen. Das Internet der Dinge (IoT – Internet of Things) ist folglich ohne Embedded Funktechnologie nicht denkbar. In diesem Artikel werden einige Aspekte des Embedded Designs funkbasierter Systeme näher erläutert, um Strategien und technische Anforderungen besser abschätzen können.

Keywords: Funk, ISM, SRD, IoT, Wireless, COTS, CUL, SOC, Industrie 4.0, Smart Factory

1 Einleitung

Das Internet der Dinge wächst. Aus Fachkreisen wurden für das Jahr 2015 15 Milliarden an das Internet angeschlossene Geräte prognostiziert. In weiteren fünf Jahren, bis 2020, sollen es gar 50 Milliarden Geräte sein. Bei aktuell 7,7 Milliarden Menschen auf der Erde benötigt nicht einmal die Hälfte jeweils fünf internetfähige Geräte um die Zahlen zu erreichen. Bei alle den Internetfähigen Systemen wie Mobiltelefone, Tabletts, iPods, Laptops und Rechner im Büro und Zuhause wird die Zahl schnell erreicht. In allen Teilen der Erde werden kaum noch Geräte verkauft die nicht mehr Internet-fähig sind, allen voran Mobile Geräte der Kommunikation, Settop-Boxen, HiFi-Anlagen oder Fernseher. Auch in der Lifestyle-Industrie sind vergleichbare Trends deutlich sichtbar. Fitnessarmbänder überwachen den Lebenszyklus und bestimmen die Essgewohnheiten. Brustgurte oder Uhren messen die Herzfrequenz oder das EKG und die Sauerstoffkonzentration des Bluts. Die Aufzeichnungen werden mit dem Mobiltelefon oder einem anderen elektronischen Gerät ausgetauscht – und in der Regel in der Cloud gespeichert.

Funktechnik ermöglicht hierbei einen hohen Schutz vor Staub und Feuchtigkeit und bietet darüber hinaus mehr Freiheitsgrade im Gerätedesign. Betrachtet man im Weitern den professionellen Bereich wie Sicherheits- und Überwachungskameras, M2M-fähige technische Systeme oder die Vielzahl der Automatisierungssysteme für die industrielle Produktion, Logistiksteuerung und Gebäudeautomation, dann wird schnell deutlich, dass Funk eine wirklich bedeutende Rolle bei eingebetteten Systemen spielt. Analysten behaupten, daß Funk die Enabler- und Schlüsseltechnologie für Industrie 4.0 ist.

1 FH Aachen - University of Applied Sciences;Lehr- und Forschungsgebiet Mechatronik und Eingebettete Systeme, Goethestr. 1, 52064 Aachen, wollert@fh-aachen.de

2 Entscheidung Funktechnologie

Das Internet der Dinge ist ohne Funktechnik nicht realisierbar. Mobile Systeme durch Kabelverbindungen in ihrer Mobilität einzuschränken macht keinen Sinn. Dieses betrifft die Systemauslegung in zweifacher Hinsicht. Auf der einen Seite ist das Energiemanagement zu betrachten. Ein günstiges Design, mit herausragenden Low-Power-Eigenschaften ermöglichen eine lange Batterielebensdauer oder einen vollständig Energie autarken Betrieb. Auf der anderen Seite spielt die Kommunikationsfähigkeit der Systeme eine bedeutende Rolle. Die Kommunikationsintensität, also Datenmenge und Häufigkeit haben einen ganz wesentlichen Einfluss auf die Lebensdauer des Energiespeichers. Die eingesetzte Funktechnologie entscheidet aber auch über die Kompatibilität zu anderen Geräten oder Systemen. Nur Systeme mit gleicher Kommunikationstechnologie können Daten untereinander austauschen. Das betrifft nicht nur den Frequenzbereich der Funktechnik sondern vielmehr den eingesetzten Protokollstack und die implementierten Geräteprofile. Gutes Systemdesign entscheidet sich dann an der spezifischen Aufgabenstellung.

2.1 Technologiekompatibilität

Kompatibilität ist eine Facette die schon fast alleine entscheidet, welche Technologie zum Einsatz kommt. Mobiltelefon und Tabletts sind aktuell die Technologietreiber. Das Nachrüsten von technologischen Schnittstellen in dieser Gerätekategorie ist in der Regel weder wirtschaftlich noch attraktiv. Der Charakter, die Haptik und das Design der Geräte sollte nicht verletzt werden, damit die Bedienakzeptanz bleibt.

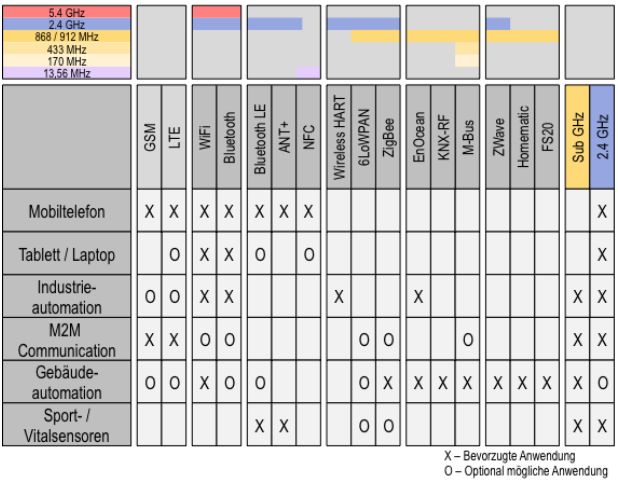


Abb.1: Spezifische Technologien ermöglichen einen differenzierten Geräteeinsatz

In der vorhergehenden Abbildung sind wesentliche relevante Technologien und ihre Anwendung in technischen Systemen tabellarisch kategorisiert. Mit einem X gekennzeichnete Felder beschrieben die typische uneingeschränkte Anwendung und die typische Verbausituation. Ein O kennzeichnet die optionale, wenn auch nicht immer anzutreffende Gerätebestückung. Ein leeres Feld kennzeichnet, dass die Technologie in dem technischen System in der Regel nicht zu finden ist.

2.2 Protokollkompatibilität

Eine weitere wesentliche Herausforderung ist der verwendete Protokollstack. Das Erzeugen von Funksignalen, deren Übertragung und das Empfangen ist nur für proprietäre Punkt-zu-Punkt Verbindungen einfach. Ist eine Kompatibilität zu anderen Systemen erforderlich, dann müssen der Verbindungsablauf wie das Anlernen von Geräten und der Verbindungsauf- und Abbau, sowie die übertragenen Nachrichten genau definiert sein. Dieses wird in den spezifischen Protokollstacks und Profilen der jeweiligen Technologien genau definiert.

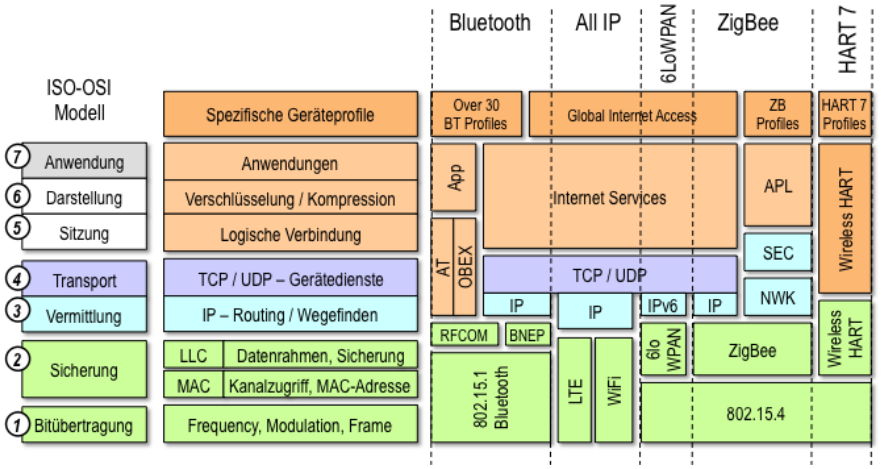


Abb.2: Der ISO-OSI-Stack macht es deutlich: Spezifische Technologien besetzen unterschiedliche Protokolllayer

Die Vorhergehende Abbildung macht es deutlich: um eine Technologie wie Bluetooth oder ZigBee zu implementieren, ist ein umfangreicher Softwareprotokollstack notwendig. Wenn nur Teile der Funktionalität des kompletten Stacks benötigt werden regeln Profile die individuelle Verwendung der einzelnen Protokollelemente. Bluetooth und ZigBee sind gute Vertreter der konfigurierbaren Protokollstacks. Mit Profilen für Anwendungsbereiche wie Automotive, Medical oder Buildingautomation sind sehr differenziert auch kleinste Applikationen möglich.

Gleichzeitig ist der Trend „All-IP“ erkennbar. Hierunter versteht man, dass alle Geräte

die Internetprotokolle IP und TCP bzw. UDP in ihrer Kommunikationsstruktur einbinden – ohne TCP/IP ist das Internet der Dinge (IoT) nicht möglich. Eine globale Sichtbarkeit von Geräten im Internet setzt eine einheitliche Struktur voraus – und das ist mit der Implementierung des TCP/IP Protokollstacks gegeben. Leider geht All-IP auch mit einer vergleichsweise hohen Komplexität und damit einem hohen Bedarf an Speicher und Rechenleistung einher. Diesem Trend kommt 6LoWPAN (IPv6 Low-energy Wireless Personal Area Network) entgegen, was eine schlanke IPv6 Anbindung über einen IEEE 802.15.4 Radio ermöglicht, darüber hinaus werden neue Versionen von „Bluetooth SMART“ auch eine vollständige IP-Unterstützung bieten. Dieses bringt IP an Geräte, bei denen sich WiFi aus Energie- und Kostengründen ausschließt.

Für den Embedded-Bereich stellt sich sehr schnell die Frage welche Technologie sich mit welchem Aufwand Implementieren lässt und welche Skalierungsmöglichkeiten bestehen. Das ist eine Frage der Embedded Chipsätze und des Systemdesigns.

3 Embedded Hardware Design

Das Design von Hochfrequenzsystemen ist nicht trivial. Auch im Zeitalter von hochintegrierten Chipsätzen ist ein profundes Knowhow im Umgang mit HF notwendig. Das zeigt sich in der geschickten Anordnung von Leiterbahnen, der Groundführung und der Spannungsversorgung sowie der Antennenanpassung. Alles zusammen ist für die Qualität eines Systems von größter Bedeutung. Ein Designfehler in der Hardware-anbindung kann schnell das Linkbudget dezimieren und damit Reichweite und Low-Power-Verhalten dramatisch verschlechtern. Darüber hinaus müssen Funksysteme zertifiziert werden und im besonderen Funksysteme mit Markenrechten wie Bluetooth®, EnOcean®, ZigBee® oder ZWave® müssen ein Zertifizierungsprogramm durchlaufen.

Um dennoch den Modulherstellern die Chance einer akzeptablen Funkimplementierung zu ermöglichen, haben sich heute vier verschiedene Vorgehensweisen etabliert.

Der erste Vertreter dieDie Königsklasse der Funksysteme ist das eigene Hard- und Softwaredesign. Dieses lohnt sich im Besonderen wenn große Stückzahlen zu erwarten sind. Zumeist wird auf kostengünstige Wireless Transceiver gesetzt, die hinsichtlich ihres spezifischen Einsatzgebietes optimiert werden. Die gesamte Software, inklusive Firmware und Protokollstacks, sowie die eigentlichen Applikation werden in dem Host-Mikrocontroller implementiert. Derartige Systeme erfordern die größte Erfahrung in der Implementierung von Funksystemen und haben gleichzeitig die größten Aufwände in der Softwareerstellung. Der Vorteil ist der uneingeschränkte Zugriff auf alle notwendigen Ressourcen. Diese Lösung bleibt Spezialisten vorbehalten.

Die zweite Skalierungsstufe sind die sogenannten NCPs. Network Co-Prozessoren bearbeiten den vollständigen Protokollstack der eingesetzten Technologie und stellen die gesamte Funktionalität des Netzwerks einem Host-Mikrocontroller zur Verfügung. Der Entwickler benötigt keine detaillierten Kenntnisse der eingesetzten Funktechnik, sondern

er kann sich auf die Programmierung und Nutzung der Anwendungsschnittstelle reduzieren. Das ermöglicht deutlich verbesserte Time-to-market-Zeiten im Vergleich zur ersten Lösung (Partioning A). Vielfach sind bei NCPs auch komplette Module verfügbar, so dass sich der Systementwickler um den HF-Teil der Anwendung und der Protokollstacks keine Gedanken machen muss.

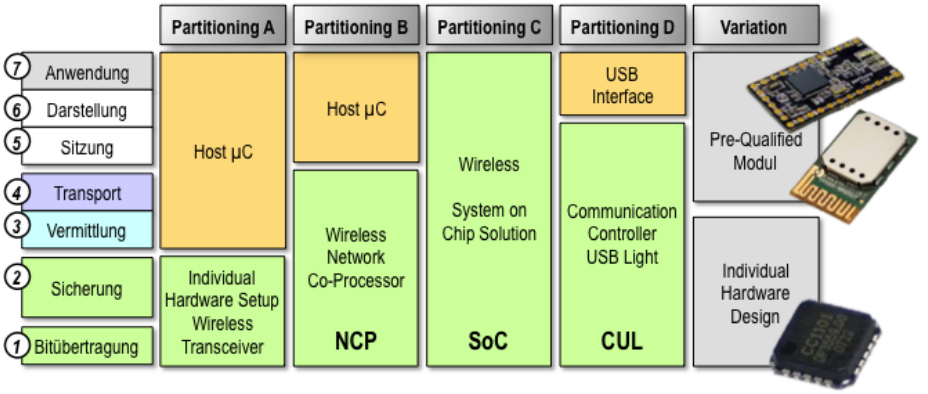


Abb.3: Von der klassischen 2-Chip-Lösung bis hin zu System-on-Chip-Lösungen sind verschiedene Realisierungsoptionen mit unterschiedlicher Komplexität möglich

Für extreme Low Power Anwendungen kommen zumeist SoCs zum Einsatz. So genannte System-on-Chip Systeme ermöglichen die größtmögliche Integration und Reduktion des Footprints. Da die Wireless-Kommunikation in der Regel die größte Herausausforderung der gesamten Anwendung ist, stellen SoCs die gesamte Kommunikation innerhalb des Chips zur Verfügung und bieten den Programmierer für seine Anwendungen eine API an und die Möglichkeit seinen Usercode auf dem SoC laufen zu lassen. Auch hier findet sich eine Vielzahl unterschiedlicher Modulhersteller, die vorqualifizierte Funkmodule in unterschiedlichsten Funktechniken und Prozessorausstattungen am Markt anbieten.

Eine spezielle Variante dieser SoCs sind die so genannten CULs. Diese Communication Controller USB Light sind spezialisierte SoCs, die über eine USB-Schnittstelle einen virtuellen seriellen Zugang zu einem 1. Level Host wie z.B. einem PC, einem Raspberry PI oder vergleichbaren Systemen zulassen. CULs werden gerne mit etablierten Betriebssystemen wie Windows, MAC OSX, Linux oder Android eingesetzt, da hier auf eine standardisierte Softwareschnittstelle gebauten werden kann. Der Softwareentwickler muss keine spezialisierten Treiber entwickeln sondern kann sich auf seine Applikation reduzieren. Mittlerweile haben sich Communities zusammen gefunden, um auf der Basis von CULs auch komplexeste Systeme aufzusetzen Als Beispiel kann hier FHEM genannte werden. Dieses Open-Source Home Automation System setzt auf den Standard Raspberry PI und bietet die spezifischen Funkschnittstellen der Gebäudeautomation fast ausschließlich über CULs ein.

Darüber hinaus bieten die führenden Chiphersteller unterschiedlichste vorgefertigte Systemlösungen an. Für Multi-Channel-Sub-GHz-Anwendungen hat Texas Instruments verschiedene Lösungen auf der Basis des CC1101 Transceiver Bausteins im Programm. Welche SoC-Lösung letztendlich eingesetzt wird hängt ganz wesentlich von der Zielrichtung der Applikation ab. Klassische Low-End-Anwendungen sind mit dem 8051-kompatiblen CC1110 zu realisieren. Durch eine hohe Integration sind einfache 8-Bit Applikationen für Low-Power-Anwendungen umzusetzen. Der CC430 SoC integriert den populären 16-Bit-Low-Power-µC MSP430 mit einem CC1101 Radio.

Bei allen CC1101-basierten Radios ist eine optionale Leistungssteigerung durch den CC1190 Rangeextender möglich. Hierbei leiden systembedingt die Low-Power Eigenschaften, ermöglicht aber eine weitreichende Skalierung der Systemschnittstelle.

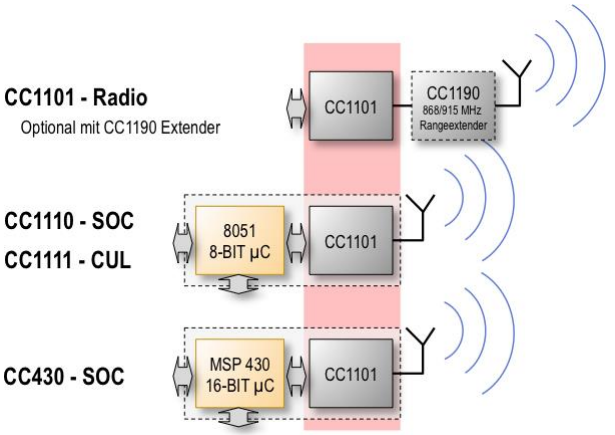


Abb. 4: TI – Sub-GHz Funkfamilie mit unterschiedlichen µC

4 Technische Low-Power Realisierung

Ein der größten technischen Herausforderungen ist das Energiemanagement bei Funksysteme. Ziel ist immer ein System, welches so gut wie keine Energie verbraucht, dabei Energieautark betrieben werden kann, eine große Reichweite hat, unbegrenzt viele Teilnehmer im Netzwerk halten kann und darüber hinaus eine hohe Datenrate bis hin zu Internet Protocol (IP)-Kompatibilität. Das ist jedoch bei weitem nicht realistisch. Mit jedem übertragenen Bit wird Energie abgestrahlt und das bedeutet, dass viel Daten auch viel Energie kosten. Das Selbe gilt für die Reichweite. Bei gleicher Empfänger-empfindlichkeit hilft nur eine große Sendeleistung. Und da zwischen Entfernung und Reichweite eine quadratische Abhängigkeit herrscht hat man nur sehr eingeschränkte Möglichkeiten Energie zu sparen.

4.1 Klassifizierung von Gerätegruppen

Sensoren und Aktoren sind die erste Gruppe der technischen Systeme. Sensoren und Aktoren können ein ausgezeichnetes Low-Power-Verhalten aufweisen, was durch einen geringen Duty-Cycle, kleine Sendeleistungen, hohe Empfindlichkeiten und optimierte Protokolle charakterisiert ist. Die Protokolle sind für die spezifische Anwendung optimiert. Wird als Gateway ein Mobiltelefon oder ein anderes Gerät dieser Kategorie eingesetzt, dann muss die Low Energy Funktechnik Bluetooth, Bluetooth LE, ANT+ oder NFC sein. Andere Techniken kommen bei kommerziellen Gateways nicht vor. Ist eine Integration mit Mobilen Consumergeräten nicht gewünscht, dann müssen spezielle Gateways in die IoT-Welt integrieren. Hier ist man offen zu allen Schnittstellen und allen Funksystemen, sie müssen nur den speziellen Anforderungen genügen.

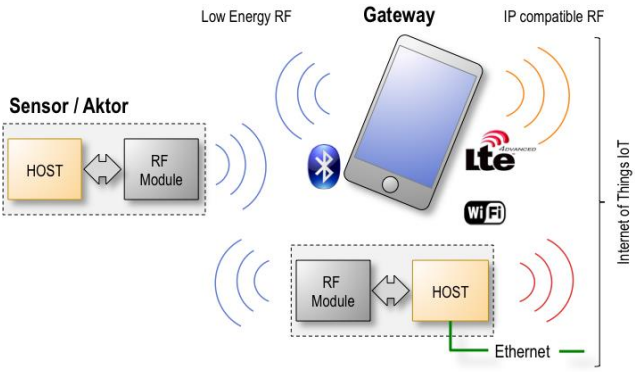


Abb.5: Das Internet der Dinge nutz IP - Ultra Low Power Systeme benötigen ein Gateway zum Anschluss an das Internet

Technologie-Gateways sind die zweite Gruppe der technischen Systeme. Diese integrieren die Low Energy Funktechniken zu der Internet Protokoll kompatiblen Welt. Gateways brauchen nicht zwingend Low-Power-Fähigkeiten besitzen, da sie letztendlich als Gateway ohnehin ständig betriebsbereit sein müssen. Mittlerweile nutzen viele Gateways standardisierte Betriebssystemplattformen wie Linux oder Android. Ganze Technologiezweige haben sich mittlerweile auf standardisierte skalierbare Embedded Systeme spezialisiert. Raspberry PI, Arduino oder Beagle Bone sind nur einige der leistungsstarken und kostengünstigen Vertreter dieser Systemgattung.

Schließlich ist als dritte Möglichkeit noch eine direkte IP-Anbindung von Sensoren denkbar. Doch auch bei 6LoWPAN, was genau diesen Bereich abdecken soll, ist der Wechsel auf eine Low-Power Funktechnik, in diesem Fall einem IEEE 802.15.4 Layer, notwendig. Letztendlich ist diese Lösung nur ein Spezialfall der vorgestellten Varianten 1 und 2 wobei das Gateway reine Routingfähigkeiten besitzen muss und damit sehr einfach standardisierbar ist.

Es ist nur zu offensichtlich, dass eine klare Trennung von Low-Power-Netzwerken zu

IP-kompatiblen Netzwerken sinnvoll ist, um allen Anforderungen gerecht zu werden. Die Wahl der Gatewaytechnologie oder des Gateways ist dann auch entscheidend für die Kompatibilität zum Rest der Welt und dem spezifischen Anwendungsfall. Im weiteren werden nun einige Technologien näher beleuchtet, die sinnvoll im Embedded Kontext Verwendung finden.

4.2 Bluetooth / Bluetooth LE

Exemplarisch soll Bluetooth als ein Vertreter potentiell geeigneter Embedded Funktechniken genannt werden. Kabellose Eingabegeräte und medizinische bzw. Sport-Sensoren gehören zu den etablieren Embedded Geräten mit der Funktechnologie Bluetooth. Seit fast 15 Jahren ist Bluetooth am Markt und hat in den letzten fünf Jahren quasi in jedes Mobiltelefon, in jeden mobilen Computer oder auch dem Auto Einzug gehalten. Bluetooth Audio-Übertragungsstrecken schaffen die Freiheit vom Kabel zur HiFi-Anlage und Bluetooth-Freisprecheinrichtungen ermöglichen eine legale Telefonie während der Autofahrt. Auch „Spielzeug“ wie Lego®s RCX oder EV3 Computer besitzen diese Schnittstelle für einen unkomplizierten kabelfreien Datenaustausch.

Mit dem Alter der Technologie hat sich auch eine entsprechende Reife ergeben. Während in der 2000er Wende die Freiheit vom Kabel noch belächelt wurde, die angedachten Szenarien für utopisch gehalten wurden und die Energieverbräuche einen sinnvollen Batteriebetrieb kaum ermöglichten, sind heute die Chiphersteller teilweise mit der 8 Generation Chipsätze auf dem Markt. Das führt zu einer erfreulichen Hochintegration, einer hohen Energieeffizienz und einer akzeptablen Interoperabilität. Mit Bluetooth LE (Low Energie) in der aktuellen Bluetooth 4 Spezifikation sind auch für einfache Anwendungen eine gute Skalierbarkeit und ein wirklich günstiges Betriebsverhalten gegeben. Diese Reife zeigt sich in den aktuellen verfügbaren SoCs, die trotz komplizierten und aufwendigen Protokollstack, eine äußerst effektive Time to Market Umsetzung ermöglichen. Neben WiFi ist Bluetooth die Connectivity-Technologie schlechthin, die zudem auch Low-Power fähig ist.

Etwa 1 Milliarde Geräte mit Bluetooth-Technologie werden pro Jahr an den Kunden gebracht. Hier von einer Nischen-Technologie zu sprechen ist deutlich untertrieben. Kaum eine andere Embedded-Funk-Technologie kann auf einen ähnlichen Erfolg verweisen. So groß wir der Markt ist, so unterschiedlich sind auch die Chipsätze und SoC-Implementierungen. Man kann feststellen, dass es Hersteller gibt, die sich ganz klar dem Massenmarkt verschrieben haben (Cambridge Silicon Radio – CSR) oder andere die auch den Embedded-Markt (Texas Instruments, Nordic Semiconductors) bedienen. Für welchen Hersteller man sich entscheidet kommt alleine auf die Zielapplikation des Embedded Gerätes an. Die Unterschiedlichen SoCs sind häufig auf spezifische Profile optimiert, so dass die Implementierung nicht ausufert sondern nur die Teile der Spezifikation implementiert sind, die auch tatsächlich benötigt werden. Die folgende Abbildung gibt einen sehr kleinen Einblick in die Bandbreite und Leistungsfähigkeit einiger verfügbaren SoCs.

Ein Trend ist klar erkennbar. Mit Bluetooth 4.1 wird Bluetooth kooperativer zur LTE Funktechnik, das seit 15 Jahren angekündigte Scatter-Net wird endlich Realität und für das Internet der Dinge wird mit dem nächsten Softwarerelease auch IPv6 zum Standard gehören. Damit ist Bluetooth aktuell wie nie. Möchte man mit einem kleinen Low-Power Gerät zu Smartphones, Tablets und Laptops kompatibel sein, dann kommt man an einer Bluetooth-Lösung kaum vorbei.

	Ti CC2540	DA14580	nRF51822	CSR8645
Hersteller	Texas Instruments	Dialog Semincond.	Nordic Semicond.	Cambridge Silicon
Technology	Bluetooth LE	Bluetooth Smart	Bluetooth / Multipr.	Bluetooth 4.0
Frequency	2.4 GHz	2.4 GHz	2.4 GHz	2.4 GHz
Processor	8051 MCU	Cortex M0	Cortex M0	80MHz RISC+DSP
Memory		32 k OTP 50 k SRAM 84 k ROM	16 k RAM 256/128k Flash	56 k RAM EEProm Interface
Data Rate	1Mbps	1 Mbps	250kbps, 1,2 Mbps	1,2,3 Mbps
TX Power	-20 .. +4 dBm	-20 .. +0 dBm	-20 .. +4 dBm	-20 .. +9 dBm
RX Sensitivity	-93 dBm	-93 dBm	-93 dBm BT LE	-92 dBm BT LE
TX current	31.6 mA @ 4 dBm 27 mA @ 0 dBm	3.4 mA @ 0 dBm 3V Supply	10.5 mA @+4 dBm 8.1 mA @ 0 dBm 3V Supply	11.8 mA Stereo high Quality
RX current	19,6 mA	3,7 mA	9,5 mA @ 1M	Slave ca. 11 mA
Shut down mode	0.4 µA	0.4 ... 1.5 µA	<1 µA	<1 µA

Abb.6: Performance einiger Bluetooth SoCs für spezielle Anwendungen

5 Fazit

Embedded Funk ist eine Notwendigkeit für IoT und Industrie 4.0. Während noch vor gut fünf Jahren über sinn und zweck von Funksensoren diskutierte, kann man heute eine umfängliche Akzeptanz feststellen. Darüber hinaus hat sich auch die Systementwicklung radikal verändert. Während im Jahrtausendwechsel für die Entwicklung von eingebetteten Funk-Systemen ein gehöriges Maß an HF-Verständnis für das Design von Systemen erforderlich war, ist man heute bei „convinience“-Produkten angekommen. Mit NCPs, SoCs und CULs stehen vorgefertigte und generische HF-Adapter zur Verfügung die problemlos in die eigenen Geräte integriert werden können. Gerade bei kleinen und mittleren Stückzahlen ist das die günstigste Möglichkeit eine Funk-Schnittstelle zu realisieren.

	WIFI - CC3000	Ti CC2530	Ti CC2540	Ti CC1101	Ti CC430
Technology	IEEE 802.11b/g	IEEE802.15.4	Bluetooth LE	Sub-GHz	Sub-GHz
Frequency	2.4 GHz ISM	2.4 GHz	2.4 GHz	300 .. 928 MHz	300 .. 928 MHz
Processor	k.A.	8051 MCU	8051 MCU	8051 MCU	MSP 430
Data Rate	1 ... 54 MBps	250 kbps	1000 kbps	1,2 .. 500 kbps	1,2 .. 500 kbps
TX Power	+18 dBm @11Mbps	-18 .. +4.5 dBm	-20 .. +4 dBm	-30 .. +10 dBm	-30 .. +12 dBm
RX Sensitivity	-88 dBm @ BER 8%	-98 dBm	-93 dBm	-111 dBm 1.2kbps	-113 dBm 1.2 kbps
TX current	11b 18dBm 11 Mbps 260 mA 11g 14 dBm 54 Mbps 200 mA	34 mA @ 4 dBm	31.6 mA @ 4 dBm 27 mA @ 0 dBm	21 mA @ 0 dBm	30 mA @ +12 dBm 15 mA @ 0 dBm
RX current	11bg / 100 mA	25 mA	19,6 mA	17 mA	14,7 mA
Shut down mode	5 µA	0.4 .. 100 µA	0.4 µA	<1 µA	<1 µA
Type	NCP: SPI/UART	SoC	SoC	SoC	SoC
Features	IEEE 802.11 b/g Embedded IPv4 Protocol Stack Simple Link ™	IEEE802.15.4 ZigBee Pro ZigBee RF4CE 6LoWPAN	Bluetooth LE compliant Software stack USB, GPIO, USART, Timer	Wireless M-Bus IEEE 802.15.4g USART, SPI, Timer, AD, AES128 Encryption	EN 300 200 IEEE 802.15.4g USART, SPI, Timer, AD, AES128 Encryption

Abb.7: Von Wifi bis Sub-GHz Technologien sind NCPs und SoCs für Embedded Systeme als generische Plattformen verfügbar

Die Herausforderung liegt in der Einbettung des Funksystems in die Systemlandschaft. Die Trends bei Industrie 4.0 und dem Internet der Dinge schreit nach All-IP für alles. Doch All-IP sagt noch nichts zu der Anwenderschnittstelle aus, die eine immer größere Bedeutung erhält. Darüber hinaus ist Ultra-Low-Power und All-IP (noch) nicht realisierbar, weil eine ständige Erreichbarkeit vorausgesetzt wird. Hier sollte verantwortungsvoll mit den verfügbaren Ressourcen umgegangen werden. Intelligente Gateway-Lösungen bringen hier Technologien, Anspruch und Wirklichkeit zusammen. Interessant bleiben zwei besondere Herausforderungen:

Interoperabilität zu Smartphone und Tablett

Sollen Handy und iPad in das Systemumfeld eingebunden werden – und zwar möglichst ohne Gateway – dann kommt man an WiFi und Bluetooth nicht vorbei. WiFi spielt eine Sonderrolle, da der System-bedingte Energiehunger bei hoher Datenrate nur mit Einschränkungen Batterie-gespeiste, oder autarke Systeme, zu lässt. Demgegenüber schafft Bluetooth SMART oder Bluetooth LE den Spagat zwischen Kompatibilität und Low Power. Intelligent aufgebaute Bluetooth LE Geräte können durchaus mit zwei AA Batterien jahrelang betrieben werden.

Integration in Systemwelten wie z.B. in der Gebäudeautomation

Embedded Systeme in der Gebäudeautomation und der Industriellen Kommunikation machen ebenfalls interoperable Systeme notwendig. Aufgrund der Ausbreitungs-eigenschaften sind hier Funksysteme im Sub-GHz-Bereich das Maß der Dinge. Mit ZigBee, KNX-RF, Wirless M-Bus oder Enocean sind verschiedenste Technologie verfügbar. Teilweise werden spezifische Radios voraus gesetzt, teilweise handelt es sich

nur um Protokollstacks die auf Standardhardware aufgesetzt wird. Hier entscheidet der Zielmarkt was die beste Lösung ist. Die Hardware-Auswahl ist im Wesentlichen durch die Verfügbarkeit von Software-Protokollstacks getrieben. Vergleichbares findet man bei Implementierungen im Health-oder Care-Markt oder bei Sport-Sensoren. Gerade beim letzten hat sich mit ANT+ auch eine 2.4 GHz Technologie etabliert.

Bei der Entscheidung für oder gegen ein Modul oder eine Technologie steht also immer der Zielmarkt im Fokus. Die Verfügbarkeit von Software-Protokollstacks ist folglich relevant für den Erfolg eines Produkts und eine gute Time-to-Market-Zeit.

Hier steckt dann auch der Kern der Problematik. Heute ist weniger die HF-Technik die Herausforderung sondern viel mehr die Kenntnis und die Möglichkeiten der Protokollstacks und die Umsetzung von Gateways und Systemlösungen. Technische Spezifikationen von führenden Technologien wie Bluetooth und ZigBee übersteigen dann schnell 1000 Seiten – und auch die wollen verstanden sein.

Literaturverzeichnis

[WOL05] Wollert, J.; Wireless Solutions? Wege in die digitale kabellose Welt, Keynote-Vortrag Wireless Technologies Kongress 12.-13.10.2005, Mannheim.

[VABW06] Vedral, A., Altrock, R., Buda, A., Wollert, J.F., "The Capability of Bluetooth for Real-time Transmission in Automation", Proceedings of the IASTED Int. Conference on Networks and Communication Systems (NCS 2006), ISBN: 0-88986-590-6, pp. 168-175, Chiang Mai, Thailand, March, 2006

[BJW07] Buda, A.; Jozenfini, B.; Wollert, J. F.: „Mobile ad-hoc Network for Home Automation based on ZigBee Application Framework, ZigBee Network Layer and nanoNET Physical Layer", 1st European ZigBee Developers' Conference, München, 18.-19. Juni 2007

[WW07] Weingart, D., Wollert, J.F., "Konvergenz kabelloser und kabelgebundener Automatisierungsnetzwerke mit Webservices und uPnP", SPS/IPC/DRIVES Conference/Wireless Corner, Nuremberg, Germany, November, 2007

[BSW08] Buda A., Schürmann V., Wollert J.F.: A Middleware Approach for IndustrialWireless Automation Networks, IASTED – Wireless and Optical Communication 2008, Quebec City, Kanada,26-28.05.2008

[BW09] Buda, A.; Wollert, J. F.: „Funktechnologien im 2,45 GHz ISM Frequenzband: Überblick und Coexistenzeigenschaften", VDI Automation 2009, 16.-17. Juni 2009 Baden-Baden, Germany, VDI-Bericht 2067, S. 383-386 – ISBN 978-3180920672

[BSW10] Buda,Schuermann,Wollert: Wireless Technologies in Factory Automation, Factory Automation, Edited by Javier Silvestre-Blanes: ISBN 978-953-307-024-7, P.29-50

[WO14] Wollert: Der Embedded Wireless-Report, Elektronik Wireless November 2014, S.18ff

[WO14_2] Wollert: Funkstandards für Sensornetzwerke, Elektronik 12/2014 S.36ff

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Entwurf und Auslegung eines Funksystems für dichte Netzwerke in der Fabrikautomatisierung

Armin Wulf[1], Wolfgang Endemann[1] und Rüdiger Kays[1]

Abstract: In der Industrieautomatisierung gibt es eine beständige Entwicklung hin zu drahtlosen Kommunikationssystemen. Flächendeckende Funkversorgung, hohe Zuverlässigkeit und geringe Verzögerungszeiten sind die wichtigsten Eigenschaften von Kommunikationstechnologien für die Fabrikautomatisierung. Mit steigender Anzahl von vernetzten Maschinenteilen und Werkstücken steigt die Teilnehmerdichte der Automatisierungsnetzwerke stark an und wird eine entscheidende Größe für den Einsatz drahtloser Technologien. In diesem Paper wird ein Mehrzellsystem mit örtlicher Frequenzwiederverwendung vorgestellt und für die Fabrikautomatisierung ausgelegt. Dabei wird die systeminhärente Interferenz unter Annahme verschiedener Kanalmodelle analysiert. Die Auswertung zeigt, dass der verwendete Empfängertyp Einfluss auf den Netzwerkentwurf hat.

Keywords: dense networks, industrial communication, inter-channel interference, network design, wireless sensor networks

1 Einleitung

Die drahtlose Kommunikation in der Industrieautomatisierung bietet viele Vorteile gegenüber kabelgebundenen Alternativen. Gesteigerte Flexibilität, die Kommunikation mit mobilen Komponenten, Nachrüstbarkeit oder Verschleißarmut sind wichtige Eigenschaften, die im Zusammenhang mit dem Übergang zu Industrie 4.0 gefordert werden. Gleichzeitig müssen die drahtlosen Technologien den strengen Anforderungen der Fabrikautomatisierung entsprechen. Gefordert werden Verzögerungszeiten von weniger als 1 ms und Paketfehlerwahrscheinlichkeiten von höchstens 10^{-9} [ZE08]. Aufgrund der Ausbreitungsbedingungen in industriellen Umgebungen stellen diese Anforderungen eine große Herausforderung für derzeit verfügbare drahtlose Technologien dar, wenn eine flächendeckende Versorgung von Produktionsbereichen erreicht werden soll. In [RSK13] konnte gezeigt werden, dass intelligente Frequenzsprung-Verfahren hierbei eine gute Performance aufweisen. Industrieautomatisierungsnetzwerke haben hohe Dichten von Funkknoten, die mit zunehmender Vernetzung von Werkstücken für Tests oder zur Koordination von Produktionsschritten weiter ansteigen. Zur Optimierung der Abdeckung empfiehlt sich eine kleinzellige Netzwerktopologie mit verteilten Access Points (APs), wobei die verfügbaren Frequenzressourcen örtlich möglichst dicht wiederverwendet werden sollen. Da diese Wiederverwendung Störungen durch Interferenz auf den gleichen bzw. benachbarten

[1] Technische Universität Dortmund, Lehrstuhl für Kommunikationstechnik , Otto-Hahn-Str. 4, 44227 Dortmund, {armin.wulf | wolfgang.endemann | ruediger.kays}@tu-dortmund.de

Kanälen verursacht, muss der Gewinn durch die Wiederverwendung höher sein, als die Einbußen durch Interferenz. Dieser Gewinn kann durch Sendeleistungsregelung und durch auf die Umgebung angepasste Platzierung von APs optimiert werden. In diesem Paper wird von einer Platzierung in quadratischen oder hexagonalen Clustern ausgegangen. Damit ist eine Vielzahl von praktisch relevanten Fällen darstellbar. Insbesondere die Anordnung der Access Points in einem quadratischen Raster stellt für typische Fabriktopologien mit Fertigungszellen eine einfach umsetzbare Lösung dar.

2 Netzwerkarchitektur

Um eine vollständige Abdeckung zu erreichen, wird im Fabrikbereich eine Anzahl Nr_{AP} von APs positioniert, die über ein kabelgebundenes Backbone miteinander vernetzt sind. Datenaustausch und Koordination der APs erfolgen über dieses Backbone. Die Netzwerkarchitektur ist in Abb. 1 dargestellt. Während die APs fest installiert sind, können die Stationen (STAs) entweder ortsfest oder frei beweglich sein. STAs können Sensoren, Aktoren oder Prozessormodule in Produkten bzw. Werkstücken sein, mit denen über das Automatisierungsnetzwerk kommuniziert werden soll. Für die physikalische Schicht des Funksystems werden nachfolgend Parameter in Anlehnung an BLE (Bluetooth Low Energy) [BT14] und WSAN (Wireless Sensor and Actor Network) [WS12] übernommen. Die Übertragung erfolgt mittels GFSK (Gaussian Frequency Shift Keying) Modulation mit einer Datenrate $R = 1$ Mbit/s, einem Bandbreiten-Zeit-Produkt $BT = 0{,}5$ und einem Modulationsindex $\eta = 0{,}5$. BLE und WSAN verwenden Frequenzsprungverfahren (frequency hopping spread spectrum, FHSS) im 2,4 GHz ISM-Band (industrial, scientific, medical band). Die Systeme nutzen

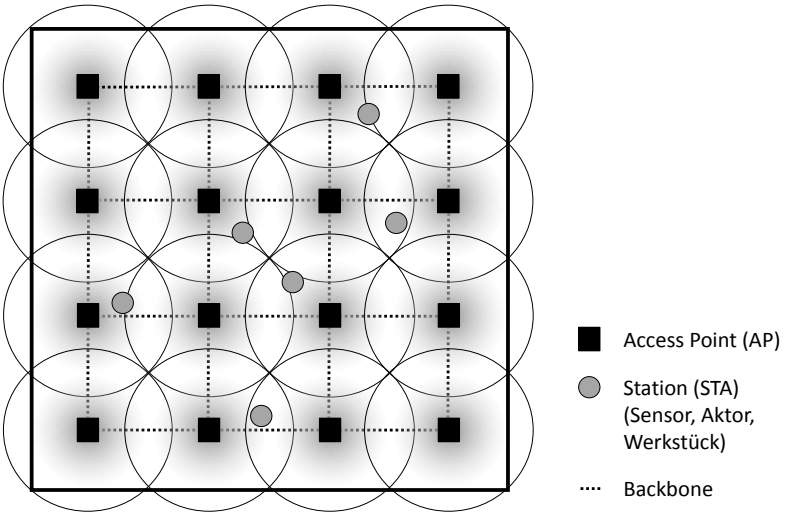


Abb. 1: Netzwerkarchitektur mit quadratischer Clusterung

unterschiedliche Kanalraster, jedoch wird bei beiden ein Kanalabstand von 2 MHz zwischen den Mittenfrequenzen der zur Übertragung verwendeten Kanäle eingehalten [BT14, WS12]. Um eine geringe Verzögerung bei der Übertragung sicherzustellen, wird – wie in der Fabrikautomatisierung üblich – ein deterministisches Kanalzugriffsverfahren verwendet, sodass durch den Medienzugriff keine zufälligen Verzögerungen verursacht werden. Zudem wird jeder STA ein eigener Uplink- und Downlinkkanal zugesichert. Als Konsequenz müssen die APs jeweils gleichzeitig auf allen Kanälen senden bzw. empfangen können, die einer Zelle zugewiesen sind. Für APs und STAs werden isotrope Antennen angenommen. Somit wird, anders als in zellularen Mobilfunknetzen, keine Sektorisierung vorgenommen. Generell entstehen durch das Clustering Netzwerkarchitekturen, die Parallelen zur klassischen Mobilfunkplanung aufweisen [ST06]. Jedoch gibt es, bedingt durch die anwendungsbezogenen Anforderungen in der Industrieautomatisierung und die daraus resultierenden, wesentlich geringeren Zellgrößen, große Unterschiede für die Auslegung. Bei umgebungsspezifischer Betrachtung der Funkausbreitung werden diese Unterschiede besonders im Hinblick auf die Interferenz zwischen Zellen mit gleicher Frequenz deutlich, und es ergeben sich, bedingt durch die Kanaleigenschaften in industrieller Umgebung [MO05], kritische Interferenzsituationen zwischen APs und STAs.

Als Frequenzband empfiehlt sich das 5 GHz ISM-Band von 5725 - 5875 MHz. Mit 150 MHz steht dort eine große zusammenhängende Bandbreite zur Verfügung, die für Frequenzsprungverfahren genutzt werden kann. Die maximale Sendeleistung von 14 dBm EIRP ist zwar kleiner als die im 2,4 GHz ISM Band maximal mögliche, jedoch müssen hier keine Koexistenzmechanismen berücksichtigt werden. Lediglich eine maximale Verweildauer von 1 s auf einer gewählten Frequenz muss eingehalten werden [ET10]. Das 5 GHz ISM Band wird im industriellen Umfeld bisher relativ schwach genutzt. Daher muss nicht mit anderen Funksystemen, beispielsweise WLAN, um die Funkressourcen konkurriert werden. Im 2,4 GHz ISM Band wäre dieses hingegen praktisch unumgänglich. Ergänzungen wie das in WSAN vorgesehene Blacklisting von ZigBee- und WLAN-Kanälen sind Schutzmechanismen, die durch die gemeinsame Nutzung des Bandes nötig werden, und deren Existenz die entstehenden Beschränkungen aufzeigt [WS12, ZE08]. Tab. 1 gibt einen Überblick über die in diesem Paper verwendeten Systemparameter.

Parameter	Wert
Modulationsverfahren	GFSK
Systembandbreite B	150 MHz
Gaußfilter $B_{3\mathrm{dB}}T$	0,5
Modulationsindex η	0,5
Kanalabstand Δf	2 MHz
Sendeleistung P_S	-20 … 10 dBm
Brutto-Datenrate R	1 Mbit/s

Tab. 1: Wichtige Systemparameter

3 Funkabdeckung mit verteilten Access Points

Ziel eines Funksystems in der Fabrikautomatisierung der Zukunft ist die Maximierung der Teilnehmerdichte unter der Bedingung, dass strikte Anforderungen an Datenrate, Paketfehlerwahrscheinlichkeit und Verzögerungszeit eingehalten werden. Ausgehend von einem festen Kommunikationszyklus und einen exklusiven Kanal für jede STA lässt sich die maximale Anzahl nutzbarer STAs Nr_{STA} im Netzwerk mit (1) aus der Anzahl der APs Nr_{AP}, der verfügbaren Bandbreite B und dem Kanalabstand Δf berechnen.

$$Nr_{\mathrm{STA}} = \frac{B}{\Delta f} \cdot \frac{1}{K} \cdot Nr_{\mathrm{AP}} \tag{1}$$

Der Wiederverwendungsfaktor K entspricht der Anzahl an benachbarten APs, die nicht dieselbe Frequenz verwenden dürfen. In zellularen Mobilfunknetzen ist K üblicherweise 3 oder 7 [GK10].

In der Fertigungstechnik ist ein Produktionsbereich meist in einzelne Fertigungszellen unterteilt, die Sensor- und Aktor-STAs enthalten. Ein Großteil der Kommunikation dieser STAs läuft innerhalb dieser Fertigungszelle ab. Aus dieser Fertigungstopologie ergibt sich auch die Aufteilung der APs für das Automatisierungsnetzwerk. Während ein quadratisches Cluster heute verbreiteten Fertigungszellen von z.B. 10 m x 10 m entspricht, ist ein hexagonales Raster im funktechnischen Sinne mit dem Ziel einer flächendeckenden Versorgung die günstigste Lösung, die man auch in Mobilfunknetzen anstrebt. Im Folgenden werden daher beide Clusterformen untersucht. Für das in Abb. 1 gegebene Beispiel einer AP-Verteilung mit quadratischen Clustern lässt sich die Anzahl Nr_{AP} der APs und damit die Anzahl Nr_{STA} der Teilnehmer pro Fläche A direkt aus dem Abstand d_{AP} der APs bestimmen.

$$Nr_{\mathrm{AP}} = \frac{A}{d_{\mathrm{AP}}^2} \tag{2}$$

Um die Anforderungen an die Paketfehlerwahrscheinlichkeit zu erfüllen, muss bei geeignetem Zellabstand eine Frequenzzuweisung durchgeführt werden, die hohe Fehlerraten durch Interferenz mit benachbarten APs vermeidet.

4 Interferenzbetrachtung bei Frequenzwiederverwendung

Der Wiederverwendungsfaktor K, der entscheidend für die Kapazität des Netzwerks ist, wird maßgeblich durch Interferenz mit benachbarten Zellen beschränkt. Um K zu bestimmen, wird zunächst angenommen, dass STAs bis zu einem maximalen Abstand r von einem AP mit diesem verbunden sein können. Daraus ergibt sich im schlechtesten Fall die in Abb. 2 dargestellte Interferenzsituation. Wenn die Sendezyklen im Netzwerk

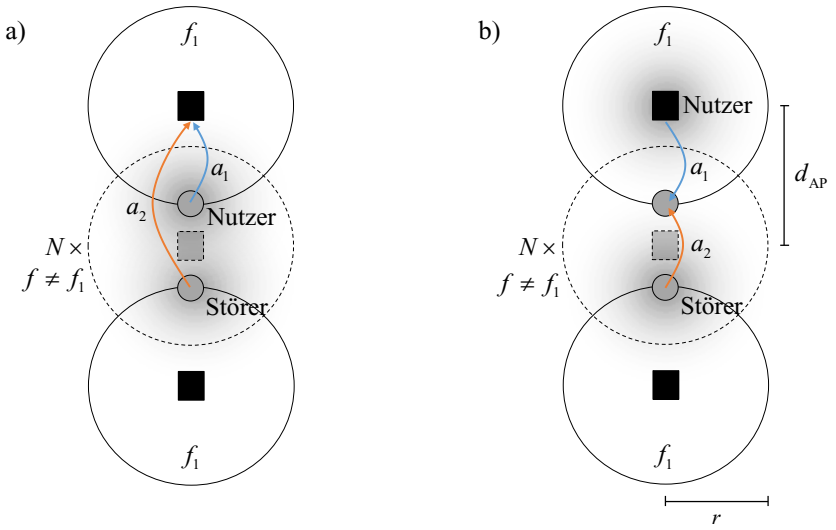


Abb. 2: Ungünstigste Interferenzsituation a) synchronisierter Sendezyklus, b) ohne Synchronisierung

synchronisiert sind, können entweder nur die APs oder nur die STAs senden. Dabei erhält man den in Abb. 2 a) aufgezeigten Fall für den Empfang an einem AP und äquivalent für den Empfang an einer STA. Besteht diese Synchronisation nicht, ergibt sich für den Empfang an einer STA die in Abb. 2 b) dargestellte Situation.

Bei LOS (Line Of Sight) Szenarien zwischen APs und STAs lässt sich die Funkfelddämpfung a durch die Freiraumausbreitung annähern. Diese ergibt sich nach (3) abhängig von Trägerfrequenz f und Distanz d zwischen Sender und Empfänger. Dabei wird angenommen, dass kein Antennengewinn erzielt wird.

$$\frac{a}{\text{dB}} = 20 \cdot \lg\left(\frac{f}{\text{Ghz}}\right) + 20 \cdot \lg\left(\frac{d}{\text{m}}\right) + 32,44 \tag{3}$$

Im Standard Bluetooth LE gelten Anforderungen an die minimale Empfindlichkeit der Receiver sowie die Auswirkungen darauf durch Interferenz auf dem gleichen oder auf benachbarten Kanälen. Die Empfindlichkeit der Empfänger muss abhängig von der verwendeten Paketlänge mindestens -70 dBm betragen bei einer Bitfehlerwahrscheinlichkeit von höchstens 10 %. Bei Interferenz auf dem gleichen Kanal darf bis zu einem Verhältnis von Signal- zu Interferenzleistung (signal to interference ratio, SIR) von 21 dB diese Empfindlichkeit um nicht mehr als 3 dB schlechter werden. Bei benachbarten Kanälen in 2 MHz Abstand darf der SIR-Wert bei -17 dB liegen [BT14]. Wählt man r so, dass eine vollständige Abdeckung entsteht, die Bereiche der APs sich

aber minimal überlappen, berechnet sich r für eine quadratische Clusterung als

$$r = \frac{\sqrt{2}}{2} \cdot d_{\mathrm{AP}} \,. \tag{4}$$

Für eine hexagonale Anordnung der APs kann r bei vollständiger Abdeckung kleiner gewählt werden. Der maximale Abstand r berechnet sich dann wie in (5) gegeben.

$$r = \frac{\sqrt{3}}{3} \cdot d_{\mathrm{AP}} \tag{5}$$

Im Fall eines synchronisierten Sendezyklus ergibt sich der SIR im schlechtesten Fall der Funkfelddämpfung gemäß (3) zu (6).

$$\frac{SIR}{\mathrm{dB}} = a_1 - a_2 = 20 \cdot \lg\left(\frac{r}{N \cdot d_{\mathrm{AP}} - r}\right) \tag{6}$$

Bei asynchronem Sendezyklus ergibt sich der SIR im schlechtesten Fall zu

$$\frac{SIR}{\mathrm{dB}} = a_1 - a_2 = 20 \cdot \lg\left(\frac{r}{N \cdot d_{\mathrm{AP}} - 2r}\right). \tag{7}$$

Wertet man die Fälle mit und ohne synchronisierten Sendezyklus für die jeweils kritische Interferenzsituation für quadratische und hexagonale Clusterung aus, ergibt sich N wie in Tab. 2 angegeben, wobei $(N + 1) \cdot d_{\mathrm{AP}}$ der Abstand zwischen zwei APs mit der gleichen Frequenz ist. Die berechneten Abstände N sind nicht vom Abstand d_{AP} der APs abhängig, jedoch wird die Zellgröße durch die maximal zulässige Sendeleistung limitiert.

Kanal	Zulässiger SIR	Sync.	Clusterung	
			Quadratisch	Hexagonal
Gleicher Kanal	21	Ja	N = 8	N = 7
		Nein	N = 9	N = 7
	10	Ja	N = 2	N = 2
		Nein	N = 3	N = 2
	6	Ja	N = 2	N = 1
		Nein	N = 2	N = 2
Nachbarkanal	-17	Ja	N = 0	N = 0
		Nein	N = 0	N = 0

Tab. 2: Abstände von APs mit gleicher Frequenz, abhängig vom zulässigem SIR und der Art der Clusterung

5 Einfluss von Interferenz auf die Receiverempfindlichkeit

Zur genaueren Betrachtung der Empfindlichkeit von FSK-Empfängern bezüglich Interferenz wurden zwei verschiedenartige inkohärente 2FSK-Empfängertypen untersucht. In Abb. 3 ist ein Blockschaltbild des äquivalenten Basisbands des Korrelationsempfängers dargestellt. Das komplexe Eingangssignal $s(t)$ wird zunächst mit den komplex konjugierten Symbolformen s_1*(t) bzw. s_2*(t) korreliert. Anschließend wird der Absolutbetrag der Korrelationen bestimmt, bevor durch Finden des Maximums der Bitwert entschieden wird. Beim Limiter-Diskriminator Empfänger (vgl. Abb. 5) erfolgt eine Amplitudenbegrenzung des komplexen Eingangssignals $s(t)$, dessen Phasenänderung anschließend durch den Diskriminator bestimmt wird. Der anschließende I&D-Block (integrate-and-dump) dient zur Rauschmittelung. Der Bitwert bestimmt sich direkt aus dem Vorzeichen der Phasenänderung. Beide Receivertypen können GFSK Signale ohne Phasensynchronisation demodulieren und sind damit gut geeignet für FHSS-Verfahren. Die Eigenschaften dieser Empfänger wurden mit zwei verschiedenen Interferenz-Modellen bei einer Bitfehlerwahrscheinlichkeit von 1 % ausgewertet. Beim SINR-Modell wird die Interferenz als weißes gaußverteiltes Rauschen modelliert, während beim Selbst-Interferenz-Modell ein weiteres GFSK Signal mit den gleichen Modulationsparametern überlagert wird. Abb. 4 zeigt die Verschlechterung γ der Receiverempfindlichkeit für beide Empfänger im SIR-Bereich von 0 – 20 dB. Während das SINR-Modell für beide Empfängertypen einen linearen Anstieg von γ für niedrige SIR zeigt, ergibt sich beim Selbst-Interferenz-Modell eine vertikale Asymptote bei SIR = 0 dB, wenn empfangenes Nutzsignal und interferierendes Signal die gleiche Leistung haben. Vergleicht man beide Modelle, zeigt sich, dass das SINR-Modell für große SIR die Verschlechterung γ der Receiverempfindlichkeit zu hoch abschätzt. Für niedrige SIR erweist sich das einfache SINR-Modell als ungeeignet. Wie in Abb. 4 dargestellt, zeigt sich unabhängig vom Empfängertyp, dass der im BLE Standard zulässige SIR für Interferenz auf dem gleichen Kanal mit 21 dB sehr vorsichtig gewählt ist. Ausgehend vom realistischeren Selbst-Interferenz-Modell mit Limiter-

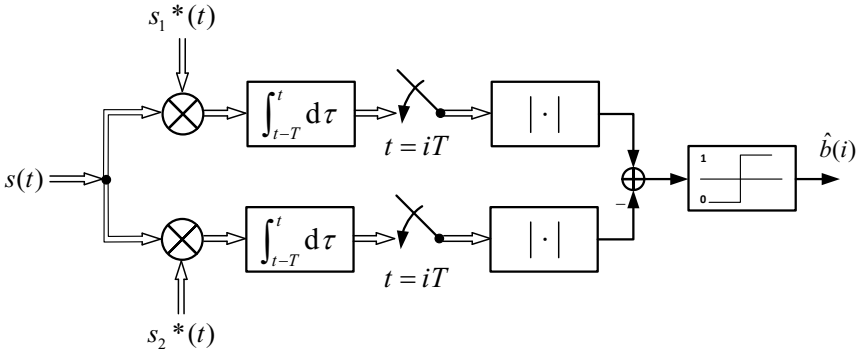


Abb. 3: Blockschaltbild des Basisbandäquivalents des Korrelationsempfängers im inkohärenten Betrieb

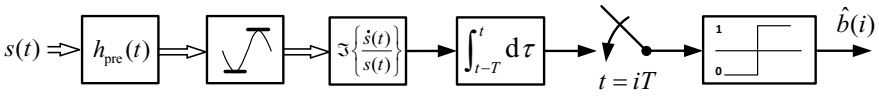


Abb. 5: Basisband-Blockschaltbild des Limiter-Diskriminator Empfängers

Diskriminator Empfänger wird die vorgegebene Verschlechterung der Empfindlichkeit um 3 dB noch bei einem SIR = 10 dB eingehalten.

Aktuelle BLE Chips haben eine Empfänger-Empfindlichkeit, die deutlich besser als die im Standard vorgegebene Mindestempfindlichkeit von -70 dBm ist. Derzeit verfügbare BLE Transceiver Chips sind beispielsweise mit Empfindlichkeiten von bis zu -97 dBm für eine Bitfehlerwahrscheinlichkeit von 0,1 % spezifiziert [TI15, DS15]. Dieses zusätzliche Linkbudget kann genutzt werden, um höhere Interferenzleistungen zuzulassen. Dies zeigt auch die in [TI15] angegebene zulässige Interferenz auf dem gleichen Kanal von 6 dB SIR bei einer Verschlechterung der Empfängerempfindlichkeit γ von weniger als 30 dB. Eine Reduktion des Linkbudgets durch Interferenz beschränkt natürlich die Reichweite der einzelnen APs, in der eine Abdeckung mit der vorgegebenen Fehlerrate eingehalten werden kann. Ein verbleibendes Linkbudget von 77 dB limitiert den Abstand der AP bei Annahme von Freiraumausbreitung mit einer

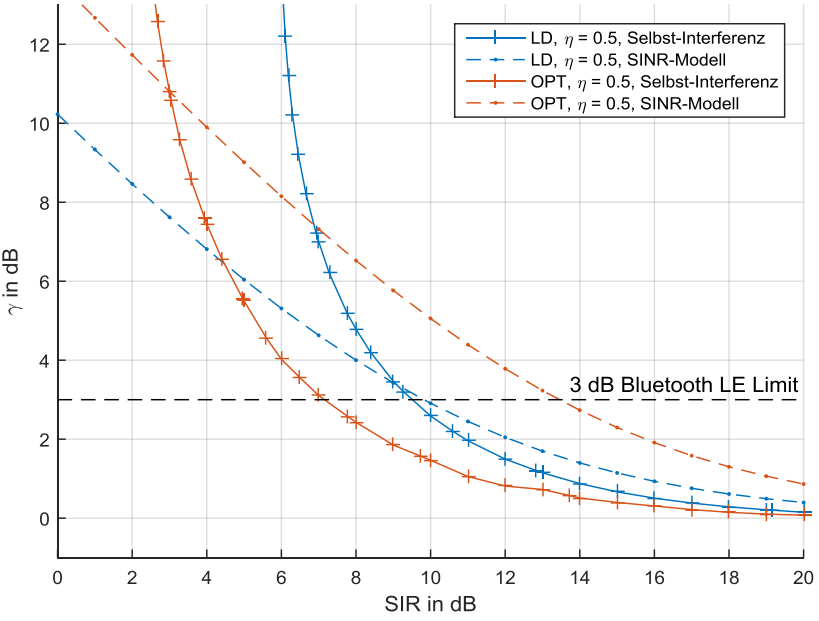


Abb. 4: Performanceeinbuße als Funktion des SIR für verschiedene GFSK Empfänger bei BER = 1 %

maximalen Sendeleistung von 10 dBm und einer Zielbitfehlerwahrscheinlichkeit von 10^{-8} und Verwendung des Limiter-Diskriminator Empfängers auf ca. 30 m bei quadratischer und auf ca. 40 m bei hexagonaler Clusterung.

6 Frequenzselektiver Kanal

Zur Berücksichtigung der Frequenzselektivität des Funkkanals in industriellen Umgebungen wird dem Pfadverlustmodell frequenzabhängiges Fast Fading überlagert. Das Fast Fading hat eine Kohärenzbandbreite, die groß gegenüber der Bandbreite einzelner Kanäle ist, sodass deren Übertragungsfunktion als frequenzunabhängig betrachtet werden können. Dem gegenüber ist das FHSS-System als Ganzes von der Frequenzselektivität betroffen. Abb. 6 zeigt die Verteilungsdichtefunktion des Übertragungsfaktors von 75 Kanälen mit 2 MHz Kanalabstand im 5 GHz ISM Band, die nach dem IEEE 802.15.4a Kanalmodell für industrielle LOS Umgebungen generiert wurden [MO05]. Als Referenz sind zusätzlich eine Rayleigh Verteilung mit einer mittleren Leistung von $\sigma^2 = 1$ und eine gefittete Rice-Verteilung mit Rice-Faktor $K = 0.22$ angegeben. Überlagert man die aus dem Fast Fading resultierende Verteilung der Übertragungsfaktoren der Einzelkanäle dem Pfadverlustmodell, das hier mit

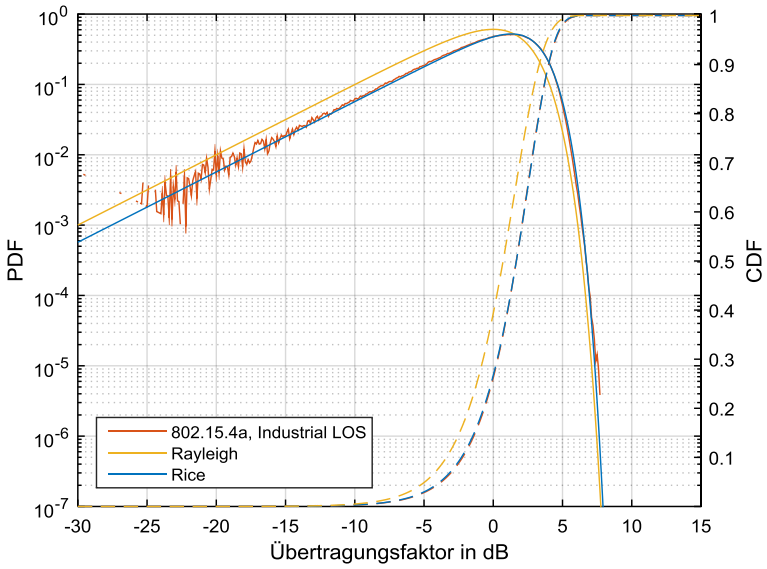


Abb. 6: Verteilungsdichtefunktion und Verteilungsfunktion des Übertragungsfaktors bei Rice, Rayleigh und 802.15a Kanalmodell

Freiraumausbreitung angenähert ist, gelten die in Tab. 2 angegebenen Zellabstände nur noch mit einer gewissen Auftrittswahrscheinlichkeit. Nur wenn der Kanal des Nutzsignals einen geringeren Pfadverlust durch das Fast Fading erleidet als die Kanäle der interferierenden Signale, ist eine Verbindung möglich. Die Fehlerwahrscheinlichkeit berechnet sich dann nach (8), wobei V_n der Übertragungsfaktor des Nutzkanals ist und V_i der Übertragungsfaktor des Kanals des $i$ten Störers ist. Die Anzahl der Störer hängt von der Frequenzverteilung ab und ergibt sich bei Vernachlässigung der mehr als $(N+1)\cdot d_{AP}$ entfernten APs für hexagonale Cluster üblicherweise zu 6 und für quadratische Cluster zu 4. ΔSIR ist der zusätzliche Abstand vom zulässigen SIR, der benötigt wird um die Zuverlässigkeit zu erhöhen.

$$P_\mathrm{err} = 1 - \prod_i \Pr\left(V_\mathrm{n} > V_i - \Delta\mathrm{SIR}\right) \tag{8}$$

Ausgehend von einem zulässigen SIR von 6 dB und synchronisiertem Sendzyklus, ergibt sich bei Erhöhung von N um 1 für die quadratische Clusterung eine SIR-Reserve von ΔSIR = 7,3 dB und bei hexagonaler Clusterung ein ΔSIR = 9,4 dB. Nimmt man an, dass Fehler nur durch die Interferenz entstehen, während das Rauschen vernachlässigt werden kann, ergibt sich für die mittlere Paketfehlerwahrscheinlichkeit 11,1 % für quadratische und 6,5 % für hexagonale Cluster. Bei zufälliger Kanalwahl sind somit bis zu 10 Kanalwechsel und Paketwiederholungen erforderlich, um die Anforderungen an Funksysteme in der Industrieautomatisierung zu erfüllen. Beim Kanalwechsel muss dabei der neue Kanal mindestens um die Kohärenzbandbreite vom alten entfernt sein, um statistische Unabhängigkeit zu erreichen. Gelingt es jedoch, durch intelligente Kanalwahl den Nutzkanal so zu wählen, dass dieser einen überdurchschnittlichen Übertragungsfaktor aufweist, ist unabhängig von der Clusterung eine SIR-Reserve von ΔSIR = 7 dB ausreichend, um die in der Fabrikautomatisierung erforderlichen Paketverlustraten zu erfüllen.

7 Zusammenfassung und Schlussfolgerung

Ausgehend davon, dass das Selektionsverhalten der FSK-Receiver unabhängig vom verwendeten Frequenzband (2,4 GHz bzw. 5 GHz) ist, zeigt Tab. 2, dass bei einem Kanalabstand von $\Delta f = 2$ MHz keine Beeinträchtigung von Nachbarkanälen zu erwarten ist. Für Übertragungen auf dem gleichen Kanal hingegen sind nach den Anforderungen des BLE Standards sehr große Abstände zwischen APs mit gleicher Frequenz von bis zu neun APs erforderlich, um den vorgegebenen SIR von 21 dB einzuhalten. Eine Verwendung von Off-The-Shelf-Empfängern, die die Receiveranforderungen des BLE Standards nur knapp erfüllen, scheidet somit aus. Eine Untersuchung der Empfindlichkeitseinbußen von sehr einfachen GFSK Empfängern zeigt jedoch, dass Limiter-Diskriminator Empfänger und inkohärenter Optimum Empfänger die geforderte maximale Verschlechterung von 3 dB bis zu einem SIR von etwa 10 dB einhalten. Mit

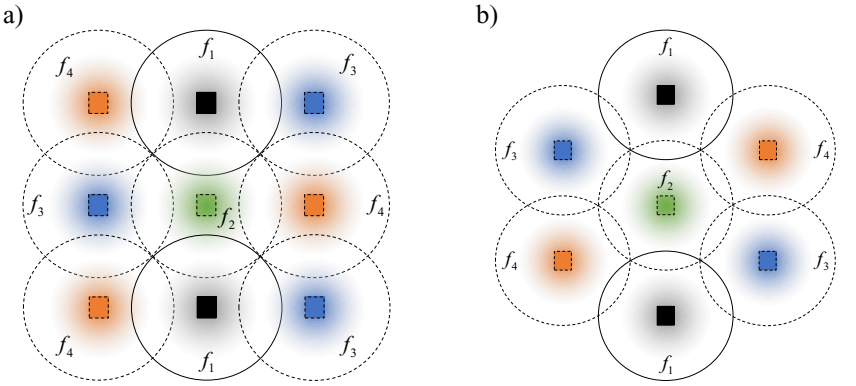


Abb. 7: Mögliche Frequenzverteilung für $N = 1$ mit a) quadratischen Clustern, b) hexagonalen Clustern

hexagonaler Clusterung ist damit, unabhängig von der Synchronisierung der Sendezyklen, ein Abstand von zwei APs zwischen den Frequenzwiederverwendungen möglich. Für den Wiederverwendungsfaktor ergibt sich damit $K = 9$. Lässt man hingegen eine größere Einbuße der Empfängerempfindlichkeit zu, wie es mit aktuellen Transceiverchips möglich wäre, lassen sich mit synchronisiertem Sendezyklus und hexagonaler Clusterung Abstände von einem AP und somit ein $K = 4$ erreichen. Dies ermöglicht große Netzwerkdichten, die durch die Abstände der APs nach wirtschaftlichen Aspekten unter den gegebenen Randbedingungen skalierbar sind. Abb. 7 zeigt beispielhaft die Frequenzverteilung für hexagonale und quadratische Cluster für $N = 1$. Für beide Cluster resultiert $K = 4$. Durch eine weitere Reduzierung des Abstands um $2/\sqrt{3}$ lässt sich der Wiederverwendungsfaktor bei hexagonaler Clusterung auf $K = 3$ reduzieren. Dadurch ergibt sich die bei Sendernetzen im Mobilfunk übliche Aufteilung.

Die vorrangegangenen Überlegungen sind für ein typisches Pfadverlustmodell mit quadratischem Abfall mit der Distanz durchgeführt worden. Dies übersteigert die Interferenz durch weit entfernte Störer, da diese in realen Umgebungen stärker als hier angenommen gedämpft würden. Dies lässt darauf schließen, dass die Auswirkungen bei hohen SIR-Grenzen zu groß abgeschätzt werden. Bei realitätsnäherer Betrachtung der Kanäle in industrieller Umgebung sind diese stark frequenzselektiv, sowie zeit- und ortsvariant [RSK13, MO05]. Das bedeutet, dass ein interferierendes Signal zufällig durch den Funkkanal begünstigt werden kann, während das Nutzsignal im ungünstigsten Fall bis zu 30 dB abgeschwächt werden kann. Durch intelligente Frequenzsprungverfahren können diese ungünstigen Kanäle gemieden werden, und bei Kenntnis des Übertragungskanals können durch intelligente Frequenzverteilung, selektiv die bevorzugten Kanäle ausgewählt werden, wodurch die Robustheit signifikant erhöht werden kann. In weiterführenden Arbeiten erfolgt eine Analyse mit genaueren

Kanalmodellen und unterschiedlichen Optimierungsverfahren für die Frequenzvergabe.

Die angestellten Untersuchungen zeigen, dass eine Frequenzmehrfachnutzung in dichten Industrieautomatisierungsnetzwerken mit aktuell verfügbarer Receiverhardware möglich ist. Für eine geeignete Zuverlässigkeit müssen jedoch sehr hohe SIR-Reserven und somit sehr geringe Wiederverwendungsfaktoren eingeplant werden. Eine statistische Untersuchung zeigt jedoch, dass bei intelligenter Frequenzwahl die Effizienz und mögliche Dichte des Netzwerks deutlich gesteigert werden kann.

Literaturverzeichnis

[ZE08] Zentralverband Elektrotechnik- und Elektronikindustrie e. V.: Koexistenz von Funksystemen in der Automatisierungstechnik, Frankfurt am Main, November 2008.

[RSK13] Reinhold, R.; Schaefer, F.-M.; Kays, R.: Performance Evaluation of an Enhanced Frequency Hopping Transceiver in 5 GHz Band for Wireless Sensor Networks. In Proc. 10th Int. Symp. on Wireless Communication Systems (ISWCS 2013), S. 1-5, Ilmenau, August 2013

[WS12] PROFIBUS Project Group PG11: WSAN Air Interface Specification, Version 1.0, PROFIBUS Nutzerorganisation e.V., Karlsruhe, Juni 2012

[BT14] Bluetooth SIG: BLUETOOTH SPECIFICATION Version 4.2, Specification Volume 6: Core System Package [Low Energy Controller volume], Dezember 2014

[ST06] Stallings, W.: Wireless Communications and Networks Second Edition. Pearson Prentice Hall, New York, 2006

[MO05] A. F. Molisch et al.: IEEE 802.15.4a channel Model - Final report, Tech. Rep. Doc. IEEE 802.15-04-0662-02-004a, 2005.

[ET10] ETSI EN 300 440-1 V1.6.1: Electromagnetic compatibility and Radio spectrum Matters (ERM); Short range devices; Radio equipment to be used in the 1 GHz to 40 GHz frequency range; Part 1: Technical characteristics and test methods. August 2010

[GK10] Ghaffar, R.; Knopp, R.: Fractional Frequency Reuse and Interference Suppression for OFDMA Networks. In Proc. 8th Int. Symp. on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), S. 273 – 277, Avignon, Juni 2010

[TI15] Texas Instruments: CC2640 SimpleLink Bluetooth Smart Wireless MCU. CC2640 Bluetooth V. 4.1 Transceiver Datasheet, Februar 2015

[DS15] Dialog Semiconductors: DA14580 Low Power Bluetooth Smart SoC. DA14580 Bluetooth V. 4.2 Transceiver Datasheet, January 2015

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Universelle Echtzeit-Ethernet Architektur zur Integration in rekonfigurierbare Automatisierungssysteme

Johannes Ax[1], Aurel Buda[2], Daniel Schneider[3], John Hartfiel[1], Lars Dürkop[4], Thorsten Jungeblut[1], Jürgen Jasperneite[3], Andreas Vedral[2] und Ulrich Rückert[1]

Abstract: Das neue Geschäftsmodell, das mit der Industrie 4.0 verbunden ist, verändert ein etabliertes Dogma der Automatisierungsbranche: Anlagen wurden auf hohen Durchsatz der immer gleichen Produktformate optimiert. Während derartige Systeme mit bekannten Maßnahmen automatisierbar sind, stellt die Variantenfertigung eher ein Automatisierungshindernis dar. Für eine individualisierte Produktion ist die Rekonfigurierbarkeit von Industrieanlagen ein wichtiger Faktor. Die Modularisierung und automatische Konfiguration sind hierbei wesentliche Triebkräfte. Im Rahmen dieser Arbeit wird eine Basisarchitektur für eine intelligente Netzwerkkomponente vorgestellt, die ohne manuelle Konfiguration in unterschiedlichen Echtzeit-Ethernet-Netzwerken eingesetzt werden kann.

Keywords: RMS, RTE, FPGA, partielle dynamische Rekonfiguration

1 Einleitung

Die Automatisierungsbranche wandelt sich zusammenhängend mit den Strömungen im Kontext der „Industrie 4.0“ [AM14, Zue10] immer weiter durch den zunehmenden Einsatz von Informations- und Kommunikationstechnologien (IKT). Es müssen daher Ansätze und Methoden entwickelt werden um die hieraus resultierenden Herausforderungen entlang der Wertschöpfungskette zu lösen:

a.) Komplexitätsanstieg durch Informations- und Kommunikationstechnologien

> Die Komplexität, der Umfang und die Reaktionsfähigkeit von Produktionssystemen verzeichnen einen starken Anstieg [BR15]. Von hoher Bedeutung sind hierbei Themen wie die Optimierung, Flexibilisierung und Beschleunigung von Produktionsprozessen. Diese Ausgangssituation löst eine zunehmende Überforderung der Menschen aus, die mit diesen Systemen, zum Beispiel für die Aufgabenfelder Programmierung, Bedienung, Instandhaltung und Service, arbeiten.

[1] CITEC, Universität Bielefeld, Kognitronik und Sensorik, Inspiration 1, 33615 Bielefeld, {jax, jhartfiel, tj, rueckert}@cit-ec.uni-bielefeld.de
[2] WAGO Kontakttechnik GmbH & Co. KG, Automation Development, Hansastr. 27, 32423 Minden, {aurel.buda, andreas.vedral}@wago.com
[3] Fraunhofer IOSB-INA, embedded systems for automation, Langenbruch 6, 32657 Lemgo, {daniel.schneider, juergen.jasperneite}@iosb-ina.fraunhofer.de
[4] inIT, Hochschule Ostwestfalen-Lippe, Langenbruch 6, 32657 Lemgo, {lars.duerkop, juergen.jasperneite}@hs-owl.de

Dieser Umstand wiederum senkt die Arbeitseffektivität der Benutzer von Automatisierungssystemen im Entwurf, im Aufbau und in der Wartung von Anlagen. Die Folge: Aufwändige Verifikation und Validierung, lange Inbetriebnahmezeiten im Anlagenhochlauf, hohe Fehleranfälligkeit durch verbleibende Hard- und Softwarefehler und damit verbunden die Gefahr von Stillstandzeiten sowie ein häufig nicht ressourcen-optimierter Betrieb der automatisierten Anlagen. Ähnliche Problemstellungen sind auch aus anderen hochtechnisierten Bereichen bekannt; die Gesellschaft für Informatik (GI) sieht hier dringenden Handlungsbedarf zur besseren Beherrschbarkeit derartig komplexer technischer Systeme [GI08].

b.) Rekonfiguration von Maschinen, Anlagen und Sensorik

In Verbindung hiermit zeigt sich ein zunehmender Trend hin zu einer hochflexiblen Variantenfertigung, welche die klassische Automatisierung an die Grenzen der Wirtschaftlichkeit bringt.

Maschinen oder Anlagen, die nie oder sehr selten umgerüstet werden, sind überwiegend auf hohen Durchsatz der immer gleichen Produktformate optimiert. Derartige Systeme können mit einer weitestgehend statischen Konfiguration betrieben werden. Produktionsprozesse mit einer hohen Variantenzahl und niedriger Losgröße haben dynamischere Anforderungen. Dies impliziert, dass Produktionsprozesse, die z.T. erhebliche Echtzeitanforderungen aufweisen können (z.B. synchrone Antriebsregelungen), mit geringen Stillstandzeiten erweitert oder umgebaut werden müssen. Benötigt werden Mechanismen, die eine Rekonfiguration mit geringem Aufwand erlauben.

Ein potenzielles Konzept ist das sogenannte „Reconfigurable Manufacturing System (RMS)“ [Ko06]. Hierbei können Module und Maschinen in einen Produktionsprozess hinzugefügt oder entfernt werden.

Ein Ansatz zur Umsetzung eines RMS ist die Einführung Service Orientierter Architekturen (SOA) in der industriellen Automatisierung. Dem Anwender wird hierdurch geholfen komplexe Prozesse und Systeme auf eine handhabbare Art zu partitionieren. Der SOA Ansatz erfordert vor allem, dass die verwendeten Feldgeräte eine semantische Beschreibung ihrer Eigenschaften und Dienste bereitstellen, um im gesamten Automatisierungssystem erkannt und nützlich eingesetzt zu werden. In [Lo13] und [Gi14] wurden Architekturen vorgestellt, um SOAs auch für Industrieanlagen mit Echtzeitanforderungen zu ermöglichen. Die Grundidee bei diesen Arbeiten ist die Kapselung der Echtzeitkommunikation in Module. Ein Modul ist eine Teilkomponente die eine spezielle Automatisierungsfunktion anbietet. Das Konzept der Module wurde auch in [Du14] aufgenommen. Die Architektur in dieser Arbeit erlaubt jedoch zusätzlich die Rekonfigurierbarkeit eines Moduls, sodass ein Hinzufügen und Entfernen einzelner Feldgeräte auch innerhalb eines Echtzeitnetzes möglich ist.

Die Echtzeitkommunikation in Automatisierungsnetzen, also auch innerhalb der hier vorgestellten Module, basiert zunehmend auf Echtzeit-Ethernet-Protokollen (RTE[5]-Protokolle). Dabei haben die letzten zehn Jahre zahlreiche RTE-Protokolle hervorgebracht, die nicht kompatibel zueinander sind. Für Hersteller von Automatisierungskomponenten sowie Analgen und Maschinenbauer ist das eine große Herausforderung. Denn Anlagenteile und Komponenten, die eigentlich die gleiche Funktion erfüllen, müssen für unterschiedliche RTE-Protokolle teils neu entworfen werden.

Aus diesem Grund wird in dieser Arbeit eine universelle Architektur vorgestellt, die unterschiedliche RTE-Protokolle unterstützt. Sie bietet neben dem Austausch von Protokollen mit harten Echtzeitanforderungen auch eine automatische Protokolldetektion. Hierdurch werden die Entwicklung von Feldgeräten und deren Integration in ein Modul vereinfacht.

2 Rekonfigurierbares Modul

Die in [Du14] vorgestellte Architektur eines Moduls erlaubt das Hinzufügen und Entfernen einzelner Teilnehmer, auch wenn zwischen diesen eine Echtzeitkommunikation stattfindet. Abbildung 1 zeigt die grundlegende Architektur eines Moduls.

Hinzugefügte Feldgeräte (IO-Devices) melden sich über einen Ad-hoc Kanal (TCP/IP-Stack) mit all ihren Eigenschaften und Diensten bei der Steuerung (PLC/SPS[6]) an. Die neu gewonnene Funktionalität wird in die semantische Beschreibung des Moduls übernommen, sodass die SPS weiterhin alle Eigenschaften und Funktionalitäten des gesamten Moduls kapselt. Über das Modul-Interface wird diese schließlich anderen Modulen zur Verfügung gestellt. Zusätzlich führt die SPS eine Autokonfiguration des RTE-Protokolls durch, wodurch der Aufwand zur Inbetriebnahme des erweiterten Moduls deutlich reduziert wird. Für weitere Details verweisen wir auf [Du14].

Als RTE-Stack (orange Blöcke in Abbildung 1) können in dieser Architektur die verschiedensten RTE-Protokolle zum Einsatz kommen. Die konkrete Auswahl hängt typischerweise von vielen Faktoren ab. Neben den technologischen Eigenschaften können benötigte Komponenten (z.B. spezielle Sensoren), die nur für ein Protokoll verfügbar sind, die Auswahl bereits bestimmen. Sehr häufig wird das Protokoll von den Auftraggebern vorgegeben (z.B. regional oder politisch bedingt). Die Unterstützung möglichst vieler RTE-Protokolle ist daher ein Wettbewerbsvorteil. Das gilt sowohl für Hersteller von Automatisierungskomponenten als auch Anlagenbauer, die diese verwenden. Wünschenswert ist eine universelle RTE-Plattform, die mit ein und derselben Hardware möglichst viele Protokolle unterstützt. Auf dieser Basis können Komponenten, unabhängig vom RTE-Protokoll, entsprechend Ihrer Semantik und Funktion entwickelt und verwendet werden.

[5] Real Time Ethernet
[6] Programmable logic controller / Speicherprogrammierbare Steuerung

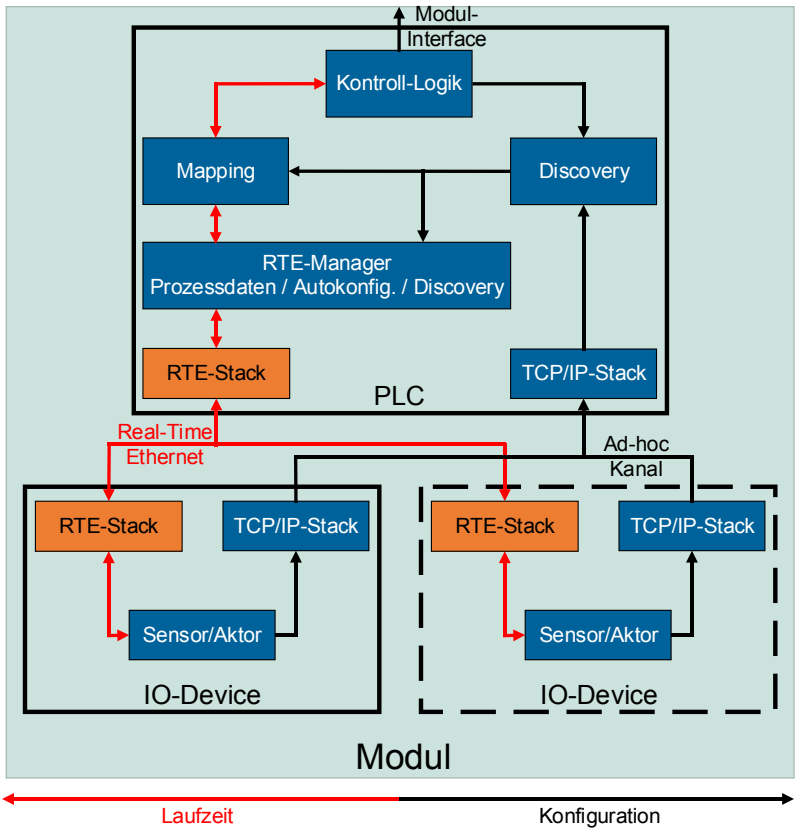


Abbildung 1 Architektur eines rekonfigurierbaren Moduls

3 Universelle RTE-Architektur

In diesem Abschnitt wird eine modulare Basisarchitektur vorgestellt, die eine dynamische Rekonfiguration des gesamten RTE-Stacks ermöglicht. Die Architektur kann zum Aufbau, sowohl von Feldgeräten als auch Steuerungen, genutzt werden. Abschnitt 3.2 geht auf die Gruppierung von RTE-Protokollen ein. In Abschnitt 3.3 werden bestehende Architekturen vorgestellt, die bereits einen Austausch von RTE-Protokollen erlauben. Abschnitt 3.3 stellt anschließend die in dieser Arbeit verwendete Systemarchitektur vor.

3.1 Kategorisierung von RTE-Protokollen

Grundsätzlich lassen sich RTE-Protokolle in drei verschiedene Kategorien (A/B/C) unterteilen [Ja05]. Die mit weicher Echtzeit bezeichneten RTEs der Kategorie A setzen direkt auf dem TCP/IP-Protokoll auf und definieren Echtzeitprotokolle in der Anwendungsschicht. Ein klassischer Vertreter ist beispielsweise Modbus-TCP [Mos06]. RTEs der Kategorie B werden ebenfalls oberhalb von Standard-Ethernet [IEE11] betrieben, sehen allerdings einen Echtzeitkanal, vorbei am TCP/IP Protokollstapel, oberhalb der MAC/LCC Schicht vor (z.B. PROFINET IO RT). Kategorie C stellt den aktuellen Stand und die höchste Leistungsklasse dar. In diese Klasse, mit sogenannten harten Echtzeitanforderungen, fallen unter anderem PROFINET IO IRT [Po10] und EtherCAT [Eth10]. Durch Modifizierungen der MAC-Schicht von Standard-Ethernet werden Zykluszeiten unterhalb von 100 µs unterstützt. Typischerweise werden hierfür ein zeitschlitzbasierter Kanalzugriff (TDMA) mit Bandbreitenreservierung und eine hochpräzise Zeitsynchronisation der Teilnehmer verwendet. Die Implementierung dieser RTEs erforderte bisher spezifische Hardwarekomponenten.

3.2 Stand der Technik

Neben dedizierten Protokoll-ICs, wie Systeme von Siemens [Si15] und Phoenix Contact [KW13], die Hardware für jeweils ein festes RTE-Protokoll der Kategorie C bieten, gibt es bereits Multiprotokoll-ICs. Diese verfügen über programmierbare Echtzeiteinheiten, die einen Austausch der unteren Protokollschichten (MAC-Schicht) ermöglichen und somit mehrere RTE-Standards der Kategorie C unterstützen. Gute Beispiele dafür sind die netX ICs der Fa. Hilscher [Hi08] und die AM335x Serie der Fa. Texas Instruments [Te15]. Auch diese Multiprotokoll-ICs stellen einen interessanten Ansatz für eine universelle RTE-Plattform dar. Zum aktuellen Zeitpunkt sind diese SoCs durch ihre geringere Leistungsfähigkeit jedoch nur eingeschränkt in der Lage, auch Zykluszeiten unter 100 µs zu bedienen. Kommende Generationen derartiger ICs können hierzu jedoch im Stande sein und könnten ebenfalls eine geeignete Plattform für eine universelle RTE-Architektur bieten.

Eine leistungsstärkere und flexiblere Plattform bieten FPGAs. Einzelprotokoll-Lösungen werden bereits von Unternehmen wie Softing und Beckhoff zusammen mit den beiden großen FPGA-Herstellern Altera [Al15] und Xilinx [Xi15] angeboten. Die neuesten FPGA-Generationen (Altera-5 Serie, Xilinx-7 Serie) bieten SOC-Systeme an, die neben der FPGA-Logik auch einen leistungsstarken ARM Cortex-A9 Dualcore enthalten, sodass die höheren Protokollschichten und Anwendungen hier ausgeführt werden können. In den Publikationen [GP06] und [GBP08] konnte nachgewiesen werden, dass die rekonfigurierbare FPGA-Logik für die harten Echtzeitanforderungen in den untersten Protokollschichten geeignet ist. Aufgrund ihrer hohen Leistungsfähigkeit der FPGA-Logik mit der kombinierten Flexibilität der SOC-Prozessoren erweisen sich SOC-FPGAs als ideale Zielarchitektur für die Anforderungen an eine universelle RTE-Architektur.

3.3 Systemarchitektur

Die im Folgenden vorgestellte Systemarchitektur bietet eine modulare Basisarchitektur für eine universelle RTE-Plattform, die ohne manuelle Konfiguration in unterschiedlichen RTE-Netzwerken eingesetzt werden kann. Als Hardwarebaustein findet hier ein moderner SoC Verwendung, der sowohl einen dedizierten Prozessor als auch FPGA-Logik integriert hat. Die FPGA-Logik bietet sich für die Umsetzung der MAC-Schicht des RTE-Protokolls an. Dies kann bei Protokollen der Kategorie A und B eine Standard-Ethernet MAC sein und bei Protokollen der Kategorie C die entsprechende echtzeitkritische MAC. Hierbei ist zu beachten, dass die verschiedenen RTE-Protokolle der Kategorie C unterschiedliche Hardwarerealisierungen erfordern. Die Vielfalt der Hardwareanforderungen wird durch die Rekonfiguration der FPGA-Logik gelöst.

Die Systemarchitektur ist in Abbildung 2 zu sehen und setzt sich aus zwei wesentlichen Teilkomponenten zusammen. Dies ist zum einen die spezifische Anwendung bzw. Funktionalität des Geräts (orange) und zum anderen der RTE-Stack (blau). Die Anwendung kann neben einer reinen Softwarelösung, die auf dem Prozessorsystem ausgeführt wird, zur Beschleunigung ebenfalls Teile der FPGA-Logik nutzen. So können für harte Echtzeitanwendungen Teile der Anwendung in Hardware abgebildet und auf dem FPGA ausgelagert werden [Ko13]. Analog zum Anwendungsfall der Rekonfiguration des RTE-Stacks, ist es mit Hilfe von partieller Rekonfiguration des FPGAs auch möglich verschiedene Teile der Anwendung auszutauschen, während die Kommunikation aufrechterhalten wird [Koe11]. Auf diesem Weg ist beispielsweise ein Firmware-Update der Anwendung während des Betriebs denkbar. Auch dieser Ansatz bietet daher schon eine eingeschränkte Möglichkeit ein Modul zu rekonfigurieren. Für eine vollständige Integration in die SOA aus [Du14] müssten semantische Beschreibungen und Dienste hinzugefügt und bereitgestellt werden (z.B. via OPC-UA).

Prototypisch wurde der Austausch von RTE-Protokollen der Kategorie C für die Standards EtherCAT und PROFINET IO (Conformance Class 3) gezeigt [Wa14]. Die spezifische MAC-Schicht wurde hierzu durch die FPGA-Logik in Hardware abgebildet. Die Konfiguration der gesamten FPGA-Logik kann durch den Prozessor während der Laufzeit durchgeführt werden. Zusätzlich wird der entsprechende SW-Stack des RTE-Protokolls auf dem Prozessorsystem ausgeführt.

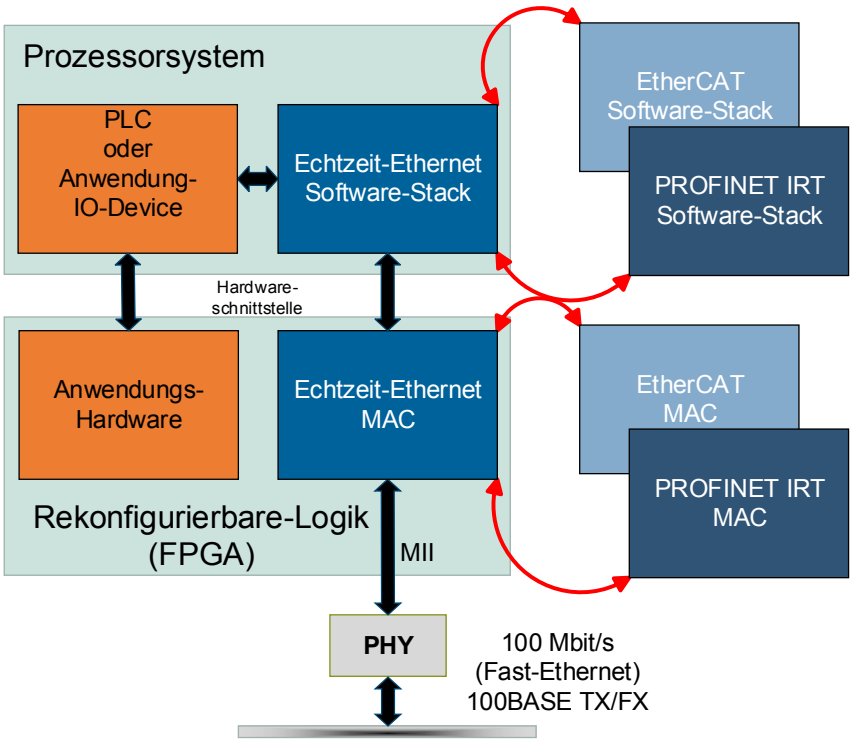


Abbildung 2 Systemarchitektur der universellen RTE-Plattform

4 Protokolldetektion von RTE-Protokollen

Um eine schnelle Integration der universellen RTE-Plattform in ein Modul zu ermöglichen, ist eine automatische Anpassung der Feldgeräte an das entsprechende RTE-Protokoll eines Moduls hilfreich. Dazu muss das verwendete Protokoll selbstständig erkannt werden.

4.1 Prinzip der Protokolldetektion

Für die Identifikation des Protokolls ist eine Auswertung der Ethernet Telegramme nötig, auf die in diesem Kapitel eingegangen wird. Der Aufbau der Telegramme ist in Abbildung 3 am Beispiel von Standard Ethernet, sowie der RTE-Standards PROFINET IO und EtherCAT dargestellt.

Es besteht aus 5 Segmenten, der Zieladresse, der Quelladresse, dem EtherType, den Nutzdaten und einem abschließenden CRC. Optional befindet sich ein VLAN TAG

zwischen Quelladresse und EtherType. Bei RTEs wird dieser Tag zur Priorisierung verwendet. Der EtherType beschreibt eine eindeutige 16 Bit ID. Anhand dieser ID kann zwischen verschiedenen Protokollen unterschieden werden. Neben Protokollen aus der Vermittlungsschicht (z.B. ARP, IPv4, IPV6, etc.) haben auch alle gängigen RTE-Standards eine eigene ID, so dass hier die eindeutige Protokollerkennung angesetzt wird. Eine vollständige Liste der EtherTypes ist unter [IEE15] zu finden.

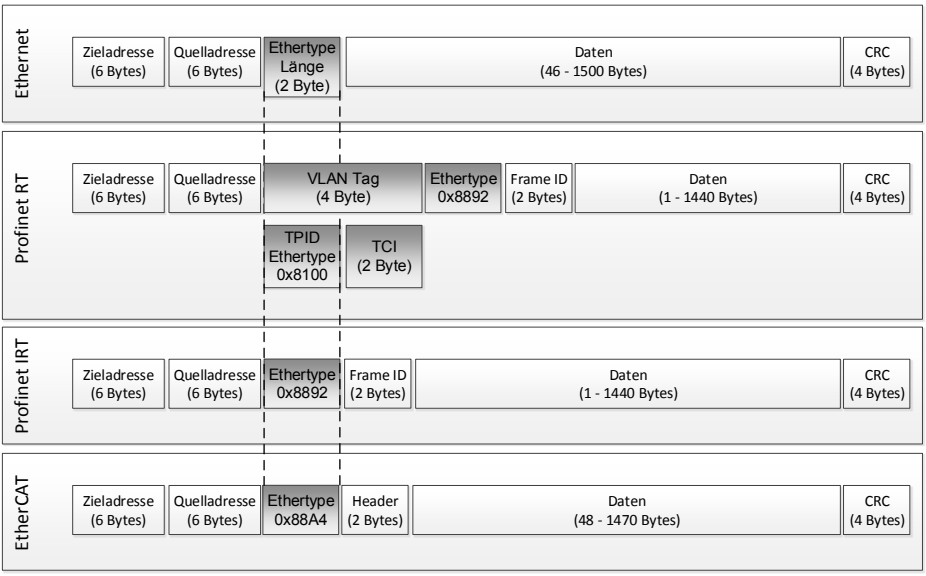


Abbildung 3 Vergleich der Telegramme von Standard Ethernet, PROFINET IO RT/IRT und EtherCAT

Nach Hinzufügen in ein Netzwerk wird der Datenverkehr analysiert und anhand des EtherTypes das Protokoll detektiert. Anschließend wird der entsprechende RTE-Stack geladen. Dies beinhaltet auch die Rekonfiguration des FPGAs mit der spezifischen MAC.

4.2 Prototypische Umsetzung

Die prototypische Umsetzung der Protokolldetektion wurde am Beispiel von EtherCAT und PROFINET IO durchgeführt. Beide Standards wurden auf dem Altera Cyclone V SX C6 SoC [Al15] abgebildet. Dieses System bietet neben einem FPGA-Bereich mit konfigurierbaren Logikelementen ein integriertes Prozessorsystem (HPS), bestehend aus zwei ARM Cortex-A9 Prozessoren.

Die Konfiguration des FPGAs erfolgt über einen Konfigurationsspeicher, auf den sowohl über externe Schnittstellen als auch über die ARM-CPU zugegriffen werden kann. Um Speicherplatz für die Unterstützung vieler verschiedener RTE-Standards einzuspa-

ren, können mit Hilfe einer partiellen Rekonfiguration auch nur Ausschnitte des FPGAs während der Laufzeit um konfiguriert werden [Wa14].

Die Protokollerkennung wird direkt bei der Systeminitialisierung durchgeführt. Dies ermöglicht eine automatische Konfiguration ohne manuelles eingreifen, nachdem die Netzwerkkomponente in das entsprechende Netzwerk hinzugefügt wurde. Abbildung 4 zeigt den Ablaufplan.

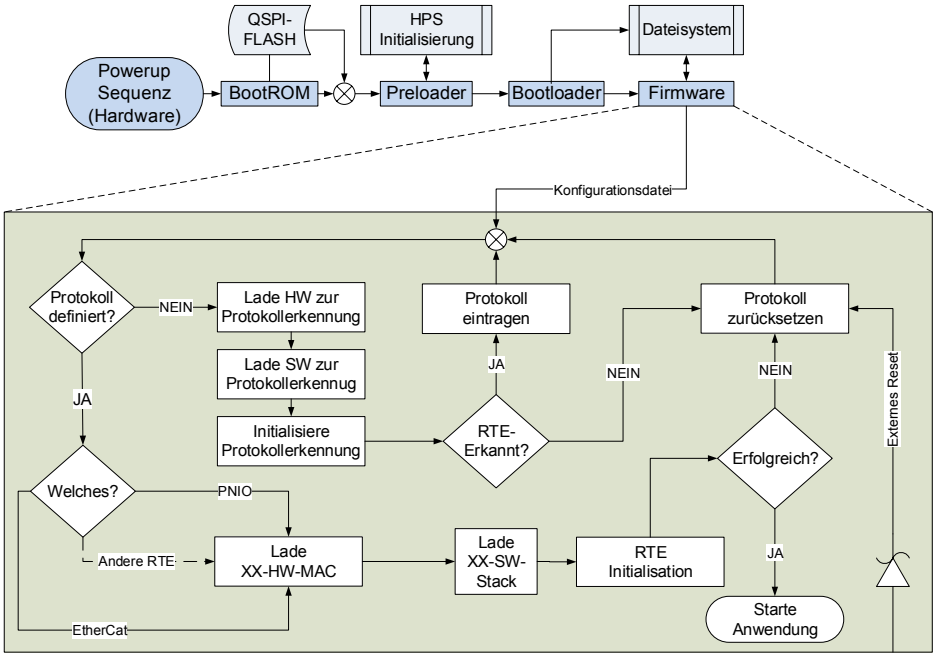


Abbildung 4: Ablauf der Systeminitialisierung mit Protokolldetektion

Nach dem Hochfahren des Systems wird aus dem BootROM heraus der Preloader gestartet, welcher sich im externen Speicher (QSPI-Flash) befindet. Im QSPI-Flash befinden sich neben dem Preloader auch der Bootloader, die zu startende Firmware (FW) sowie ein Dateisystem. Im Dateisystem liegen HW-Images der PROFINET IO und Ether-CAT Cores, sowie persistent speicherbare Konfigurationsdaten. Nachdem der Preloader das Prozessorsystem (HPS) initialisiert hat, wird der Bootloader gestartet, welcher wiederum die HPS FW startet. Die Initialisierung des FPGAs und das Laden der entsprechenden SW-Stacks werden von der FW durchgeführt.

Nach einem Start liest die FW die gespeicherten Konfigurationsdaten und wertet aus, ob bereits ein RTE-Protokoll konfiguriert ist. Ist dies der Fall, wird dessen HW-Image in den FPGA geladen und der passende SW-Stack samt Anwendung gestartet. Ist noch kein Protokoll konfiguriert wird ein spezielles HW-Image in den FPGA geladen, das zur Detektion des RTE-Protokolls dient. Hierbei handelt es sich um einen normalen Ether-

net-Switch, der zusätzlich Informationen über die EtherTypes von Telegrammen, die den Switch durchlaufen, an die FW weiterleitet. Detektiert die FW danach EtherTypes eines RTE-Protokolls, in diesem Fall 0x8892 für PROFINET IO oder 0x88A4 für EtherCAT, werden das HW-Image, sowie SW-Stack samt Anwendung für das detektierte Protokoll geladen. Eine weitere Protokolldetektion im laufenden Betrieb ist in der aktuellen Ausbaustufe nicht geplant.

Wie sich das Gerät im Falle eines Neustarts verhalten soll ist konfigurierbar. Es kann eingestellt werden, dass die Protokolldetektion nach jedem Neustart gestartet werden soll, oder dass das erste detektierte Protokoll auch nach Neustarts sofort angewendet wird. Ferner ist es möglich ganz ohne Protokolldetektion das RTE-Protokoll vorzugeben. Diese Einstellungen können jederzeit zurückgesetzt oder modifiziert werden.

5 Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurde die Architektur einer universellen RTE-Plattform vorgestellt. Hiermit können Hersteller von Komponenten unabhängig vom RTE-Protokoll Feldgeräte oder auch Steuerungen entwickeln. Anlagen- und Maschinenbauer können mit diesen Komponenten Module nach [Du14] für unterschiedliche RTEs aufbauen und schneller neue Märkte schneller erschließen.

Insbesondere die gezeigte Protokolldetektion und automatische Netzwerkkonfiguration eignen sich für eine Integration in serviceorientierte und selbstkonfigurierende Automatisierungssysteme. Bestehende „Plug and Produce“ Ansätze können hierdurch sinnvoll erweitert werden.

Die vorgestellten Konzepte sind für beliebige RTE-Protokolle anwendbar. Validiert wurden Sie am Beispiel von EtherCAT und PROFINET IO.

Für eine vollständige Integration in Service Orientierte Architekturen werden zukünftig semantische Beschreibungen und die Abbildung von Diensten oberhalb der Eigenschaften der RTE-Protokolle betrachtet. Erste Ansätze sind in [Du14] beschrieben. Die Standardisierung semantischer Beschreibungen ist aktuell Gegenstand von Industrie 4.0 Arbeitskreisen.

6 Danksagungen

Dieses Forschungs- und Entwicklungsprojekt wird mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) im Rahmen des Spitzenclusters “Intelligente Technische Systeme OstWestfalenLippe“ (it’s OWL) gefördert und vom Projektträger Karlsruhe (PTKA) betreut. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Literaturverzeichnis

[Al15] Altera: Industrial Automation. 2015. Online: https://www.altera.com/solutions/industry/industrial/applications/automation/industrial-networking.html. [Zugriff am 18 03 2015].

[Al15] Altera: SoC FPGA ARM Cortex-A9 MPCore Processor Advance Information Brief. 2012. Online: https://www.altera.co.jp/ja_JP/pdfs/literature/hb/soc-fpga/aib-01020-soc-fpga-cortex-a9-processor.pdf. [Zugriff am 18 03 2015].

[AM14] AMA Fachverband für Sensorik e.V.: „Sensor-Trends 2014", http://www.ama-sensorik.de/verband/publikationen/, 2014

[BR15] B&R: „Auf der Überholspur mit reACTION Technology“, http://www.brautomation.com/de/produkte/innovations2015/ultraschnellereaktionen/, 13.01.2015

[Du14] Durkop, L., Trsek, H., Otto, J., & Jasperneite, J.: A field level architecture for reconfigurable real-time automation systems. In Factory Communication Systems (WFCS), 2014 10th IEEE Workshop on (pp. 1-10). IEEE.

[Eth10] EtherCAT Technology Group: EtheCAT Communication. 2010.

[GBP08] Griese, B.; Brinkmann, A.; Porrmann, M.: SelfS – A Real-Time Protocol for Virtual Ring Topologies.16th International Workshop on Parallel and Distributed Real-Time Systems (WPDRTS'08), Miami, Florida, USA, 2008.

[GI08] Gesellschaft für Informatik (GI): 38. Jahrestagung „INFORMATIK 2008 – Beherrschbare Systeme – dank Informatik (http://www.informatik2008.de/371.html)

[Gi14] Girbea, A.; Suciu, C.; Nechifor, S.; Sisak, F.: Design and Implementation of a Service-Oriented Architecture for the Optimization of Industrial Applications. IEEE Transactions on Industrial Informatics, vol. 10, pp. 185–196, 2014.

[GP06] Griese, B.; Porrmann, M.: A Reconfigurable Ethernet Switch for Self-Optimizing Communication Systems. IFIP Conference on Biologically Inspired Cooperative Computing (BICC 2006), pp. 115-124, Santiago de Chile, Chile, 2006.

[Hi08] Hilscher: Technical Data Reference Guide netX 500/100 Edition 1.3. 2008.

[IEE11] IEEE: IEEE Standard for Information Technology – Telecommunications and Informationexchange between Systems – Local and Metropolitan Area Networks – Specific Requirements: Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method, 2011.

[IEE15] IEEE Online: http://standards.ieee.org/develop/regauth/ethertype/eth.txt. [Zugriff am 16.3.2015].

[Ja05] Jasperneite, J.: Echtzeit-Ethernet im Überblick. *Automatisierungstechnische Praxis,* pp. 29-34, März 2005.

[Ko06] Koren, Y.: “General RMS Characteristics. Comparison with Dedicated and Flexible Systems”, in Reconfigurable Manufacturing Systems and Transformable Factories. pp. 27–45. Springer, 2006.

[Ko13] Korf, S; Sievers, G.; Ax, J.; Cozzi D.; Jungeblut, J.; Hagemeyer, J.;Porrmann, M. Rückert, U.: Dynamisch rekonfigurierbare Hardware als Basistechnologie für intelligente technische Systeme. Wissenschaftsforum 2013 Intelligente Technische Systeme, 2013.

[Koe11] Koester, M.; Luk, W.; Hagemeyer, J.; Porrmann M.; Rückert, U.: Design Optimizations for Tiled Partially Reconfigurable Systems. Very Large Scale Integration (VLSI) Systems, IEEE Transactions on, pp. 1048-1061, 2011.

[KW13] KW-Software GmbH: TPS-1 Single Chip Interface Solution for PROFINET Devices Version 1.2.9. 2013.

[Lo13] Loskyll, M.: Entwicklung einer Methodik zur dynamischen kontextbasierten Orchestrierung semantischer Feldgerätefunktionalitäten, Technische Universität Kaiserslautern, 2013, Dissertation.

[Mos06] Modbus-IDA: MODBUS Messaging on TCP/IP Implementation Guide. 2006.

[Po10] Popp, M.: Das PROFINET IO-Buch. ISBN 978-3-8007-3274-6, VDE Verlag, 2010.

[Si10] Siemens AG: ERTEC 400 - Enhanced Real-Time Ethernet Controller Version 1.2.2. 2010.

[Te15] Texas Instruments: PROFINET on TI's Sitara processors. 2015

[Wa14] Walter, M.; Ax, J.; Buda, A. Nußbaum, K.; Hartfiel, J.; Jungeblut T.; Mario, P.: Dynamische Rekonfiguration von Echtzeit-Ethernet-Standards mit harten Echtzeitanforderungen. KommA 2014, Lemgo, Deutschland, 2014.

[Xi15] Xilinx: Industrial Automation. 2015. Online: http://www.xilinx.com/applications/industrial/industrial-networking.html. [Zugriff am 18 03 2015].

[Zue10] Zuehlke, D.: SmartFactory – Towards a factory of things. In: Annular Reviews in Control 34 (2010). S. 129-138, Elsevier Ltd., 28.03.2010

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Eine semantische Klassifikation von Systeminteraktionen

Johannes Reich[1]

Abstract: Es wird eine Klassifikation von Systeminteraktionen vorgestellt, die sich auf den Verarbeitungsmodus von Sender und Empfänger entlang dreier Dimensionen bezieht: Asynchron/synchron, deterministisch/nichtdeterministisch und zustandslos/zustandsbehaftet. Wegen der Vollständigkeit dieser Klassifikation lassen sich alle wohldefinierten Interaktionen in diese 8 Kategorien einordnen. Den Kategorien lassen sich unterschiedliche charakteristische syntaktische Mittel zuordnen.

Aus der Klassifikation folgt eine eindeutige Unterscheidung zwischen vertikaler und horizontaler Interaktion. Damit lässt sich zum einen im Bereich der Architektur- und Komponentenmodelle der formale Nachweis von Softwareschichten führen und zum anderen die Forderung motivieren, dass Komponentenmodelle für netzwerkartig operierende Applikationen Protokolldeklarationen unterstützen sollten.

Als Beispiel der Anwendung der Klassifikation werden die beiden sogenannten Architekturstile SOA und REST herangezogen. Es wird festgestellt, dass deren Grundbegriff des ”Service” nicht hinreichend klar definiert ist, dass er eine eindeutige Einordnung in die vorgestellte Klassifikation erlauben würde.

Abstract: A classification of system interactions is presented that is based on the processing mode of sender and receiver along the three dimensions: asynchronous/synchronous, deterministic/non-deterministic, and statless/statefull. Because of its completeness, every interaction can be attributed to one of the 8 categories. Different syntactical means can be attributed to each interaction class.

The classification offers a clear distinction between vertical and horizontal interaction. This allows on the one hand the formal proof of software layering in the area of architecture and component models and on the other hand motivates the requirement for component models for network-like operating applications to provide means for protocol declarations.

With respect to the so called architectural styles of SOA and REST, it is noticed that their basic notion of a ”service” is not sufficiently well defined to allow for an unambiguous attribution within the proposed classification.

Keywords: Interaktionen, synchron, asynchron, deterministisch, nicht-deterministisch, zustandsbehaftet

1 Einführung

Computersysteme interagieren auf vielfältige Weise. Dabei sind nicht alle Interaktionen von derselben Qualität, sondern es scheint verschiedene Interaktionsklassen zu geben.

Prominente Beispiele von Versuchen, Interaktionen im Bereich der Informatik zu klassifizieren sind etwa die ”Transaction Pattern” der UN/CEFACT Modelling Methodology

[1] SAP SE, Industry Standards & Open Source, Dietmar-Hopp-Allee 16, 69190 Walldorf, johannes.reich@sap.com

[UMM03] oder RosettaNet [RNI01], die durch die Theorie der Sprechakte motiviert wurden [Au62, Se69], oder die ”Transmission Primitives” [WSD01] bzw. ”Message Exchange Pattern” [WSD07a, WSD07b] aus der Welt der Web Services. Für beide Ansätze lässt sich mit Hilfe von Zustandsautomaten recht einfach zeigen, dass sie nicht gut geeignet sind, die Semantik von prozesshaften Interaktionen zu erfassen [Re08].

Insbesondere in Architekturmodellen können diese unklaren Klassifizierungen zu Unklarheiten bis hin zur Unbrauchbarkeit führen. Spätestens seit dem bekannten Open System Interconnection Modell [ITU94], wird zwischen vertikaler Interaktion zwischen Softwareschichten und der horizontaler Interaktion innerhalb einer Softwareschicht unterschieden - ohne dass klar gestellt wurde, wie sich diese beiden Interaktionsklassen formal - also genau und eindeutig - unterscheiden ließen. So redet das angesprochene OSI-Modell bezogen auf die vertikale Interaktion zutreffend von einer Verwendungsbeziehung, und bei der horizontalen Interaktion von Protokollen. Das Referenzarchitekturmodell Industrie 4.0 ([I4015], S. 43) hingegen spricht davon, dass innerhalb der Schichten eine ”hohe Kohäsion” und zwischen den Schichten eine ”lose Kopplung” herrschen soll.

Tatsächlich stellen Juha Savolainen and Varvana Myllärniemi in [SM09] fest, dass ”there is very little actual research done on layered architectures”. Diese Unklarheiten stellen aus meiner Sicht mittlerweile einen deutlichen Hemmschuh bei der technologischen Weiterentwicklung von netzwerkartig interagierenden Applikationen dar, dessen Abdruck sich auch in den Mühen, diese Weiterentwicklung systematisch durch Standardisierung zu fördern, niederschlägt, etwa im Bereich Industrie 4.0 oder dem Internet der Dinge [KLW11, BM12]. So wird im schon zitierten Ergebnisbericht der Plattform Industrie 4.0 ([I4015], S. 37) das Ziel formuliert, für Industrie 4.0-Szenarien eine formale, Computerverarbeitbare Form der Beschreibung zu entwickeln - Aber: Wie soll dies möglich sein, wenn die Unterschiede zwischen vertikaler und horizontaler Interaktion nicht formal - also genau und eindeutig - gefasst werden? Ziel dieses Artikels ist, eine einfache Klassifikation von Interaktionen aufzuzeigen, die insbesondere die beiden Klassen der vertikalen und der horizontalen Interaktion umfasst.

Ein weiterer sehr wichtiger Anwendungsfall einer Interaktionsklassifikation sind Komponentenmodelle. Mit einer vollständigen Klassifikation erhalten wir ein Kriterium, nach dem sich beurteilen lässt, ob Komponentenmodelle in diesem Maße tatsächlich vollständig sind.

Da die angesprochenen Unklarheiten meiner Ansicht nach auf einer unklaren Auffassung von der Bedeutung ausgetauschter Informationen beruhen, beschäftige ich mich im ersten Abschnitt mit der Frage, wie Informationen zu ihrer Bedeutung kommen. In den weiteren beiden Abschnitten wird die Klassifikation der Interaktionen vorgestellt. In Abschnitt 5 diskutiere ich die sogenannten Architekturstile der ”Service Orientierung” (SOA) und des ”Representational State Transfers” (REST). Der letzte Abschnitt fasst die wesentlichen Ergebnisse noch einmal zusammen.

2 Die Bedeutung von Informationen

Claude Shannon hat schon in seinem Artikel "A Mathematical Theory of Communication" ([Sh48]) darauf aufmerksam gemacht, dass die Betrachtung von Kommunikation als Informationsübertragung den Betrachter von der Bedeutung des Geschehens absehen lässt:

> "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem."

Stattdessen wird die Betrachtung der Kommunikation zunächst auf die Frage der Reproduzierbarkeit von Zuständen reduziert, eben auf die Übertragung von "Information"[2].

Die Fragen der Semantik, also der Bedeutung der übertragenen Informationen ist damit natürlich nicht irrelevant geworden, nur weil sie durch die Theorie der Informationsübertragung ignoriert wird. Ganz im Gegenteil hat sich herausgestellt, dass sich das Problem der Informationsübertragung eher bescheiden ausnimmt gegenüber dem Problem der Verständigung.

Meiner Ansicht nach hat Claude Shannon mit seiner Trennung zwischen der Übertragung und der Verarbeitung von Informationen auch für ein adäquates Verständnis der Bedeutung von Informationen die richtige Grundlage gelegt. Eine sehr interessante und gleichwohl naheliegende Folgerung aus der scharfsinnigen Idee von Claude Shannon ist nämlich, dass sich Bedeutung materiell, also in einem physikalischen Sinn, gar nicht übertragen lässt. Stattdessen erhält Information Bedeutung schlicht durch ihre Verarbeitung, man könnte auch sagen, durch ihre Interpretation. Bedeutung wird nicht transportiert, sondern erteilt. Informationen, die nicht verarbeitet werden, sind bedeutungslos.

Die dieser Betrachtung zugrunde liegende Vorstellung von Semantik ist die einer Abbildung, der Interpretationsfunktion. Eine solche Vorstellung von Semantik kommt auch in vielen anderen wichtigen Bereichen der Informatik erfolgreich zur Anwendung. Die formale Semantik imperativer Programmiersprachen wird im Wesentlichen so verstanden. Dana Scott und Christopher Strachey ([SS71]) beschreiben etwa die Semantik eines Computerprogramms durch die Abbildung der Operationsterme auf Funktionen im mathematischen Sinne (die ihrerseits wieder entsprechende Abbildungseigenschaften von der Ein- auf die Ausgabe aufweisen), während Gordon Plotkin ([Pl81]) die Semantik eines Programms zusammen mit seinen Daten durch eine Abbildung auf ein Transitionssystem erklärt und damit ebenfalls auf die Operationen fokussiert, die ein System ausführen kann.

Als Beispiel sei die Zeichenfolge 'y=Sinus(x)' angegeben, die bei der Interpretation des Programmtextes auf ein System abgebildet wird. Dieses System bildet den Wert des Zu-

[2] Im weiteren verwende ich die Terme "Daten" und "Informationen" synonym, ebenso wie die Terme "Bedeutung" und "Semantik".

stands, der durch den Namen 'x' repräsentiert wird, in einem Zeitschritt auf einen Wert des Zustands, der durch den Namen 'y' repräsentiert wird, derart ab, dass zwischen beiden Zustandswerten eine mathematische Beziehung besteht, die der Funktionsname 'Sinus' suggeriert. Eine Bedeutung des Zustandswertes von 'x' (seine Belegung) wäre dann der anschließend errechnete Zustandswert von 'y'.

Auch in der formalen Logik wird die Semantik von Termen und Ausdrücken über eine Interpretationsfunktion erklärt, die diese Ausdrücke in der Regel rekursiv auf die Begriffe einer Metasprache abbildet. Auch in der formalen Logik ist zur Auswertung von Ausdrücken, die Variablen enthalten, eine Belegung dieser Variablen mit Werten notwendig.

Mit dieser Auffassung ist konsistent, dass allen Techniken, die die Verarbeitung von Informationen strukturieren sollen, ein "semantischer Gehalt" zugemessen wird. Das gilt sowohl für Typsysteme, in denen sich Datentypen als Paare von Daten- und Operationenmengen auffassen lassen, wie auch für Abstrakte Datentypen oder sogenannter "Objekte", die einer Menge von Daten eine feste Anzahl von Operationen zuordnen.

Und es ist ebenfalls konsistent zu dieser Auffassung, dass eines der ältesten, auf Gottfried W. Leibniz zurückgehende semantische Prinzip, dass zwei Ausdrücke die gleiche Bedeutung haben, wenn sie gegenseitig ersetzbar sind (salva veritate) mit Barbara H. Liskovs und Jeannette M. Wings "Substitutional Principle" ([LW94, LW01]) Eingang in die Informatik erhielt.

Aus dieser Auffassung leiten sich unmittelbar einige Folgerung ab. Zum einen macht diese Auffassung von Semantik keinen Unterschied zwischen einer menschlichen (oder allgemein biologischen) und einer technischen Verarbeitung. Ob mir mein Geld aufgrund meiner Eingabe am Geldautomat ausgezahlt wird oder von einem Bankangestellten ist bezüglich der Abbildung: Mein Wunsch $\rightarrow$ Geldinhalt meines Portmonaies, unerheblich. Gleichzeitig unterstreicht sie die Vielfalt unserer Bedeutungen. Sie inkludiert etwa die Fregesche Vorstellung von der Bedeutung einer Aussage als ihrem Wahrheitswert [Fr92]: Sobald sich eine Aussage auf die Wahrheitswerte "wahr" oder "falsch" abbilden lässt, ist dieser Wahrheitswert eine mögliche Bedeutung dieses Satzes. Und nicht zuletzt hat Bedeutung immer lokalen Charakter, im Gegensatz etwa zur These Hilary W. Putnams [Pu75][3].

[3] Über dieses offensichtlich auch philosophisch sehr interessante Thema ließe sich sicherlich noch deutlich mehr sagen. An dieser Stelle seien mir zwei Bemerkungen zur Rolle unseres Verstandes und der Komplementarität von Verstand und Emotionen als Fußnote erlaubt.

Selber Bedeutung erteilen und verstehen, welche Bedeutung jemand drittes erteilt hat, sind zweierlei. Mit der vorgestellten Bedeutung von Bedeutung lässt sich unser Verstand, als das menschliche Organ zur Erfassung von Bedeutung auffassen, der damit das Individuum aus seiner existenziellen Isolation treten lässt.

Fasst man Emotionen als "Zustände der Interpretation" auf, also als innere Zustände, die unsere Verarbeitung von Informationen ganz wesentlich modulieren (was sie zweifelsfrei tun), dann sind Verstand und Emotionen untrennbar miteinander verbunden. Für eine solche Sicht spricht auch die enge Verbindung von Verstand und Emotionalität in den pathologischen Störungen unseres Denkapparates, den Psychosen, die sich stärker auf die Emotionen als affektive Psychosen oder auf den Verstand selbst, als kognitive Psychosen beziehen können.

3 Klassifikation von Interaktionen entlang dreier Dimensionen

Die Verarbeitung von Informationen wird durch die Theorie der Berechenbarkeit beschrieben, deren Grundlagen von Alan Turing ([Tu36]) mit der Turingmaschine, Alonso Church ([Ch36]) mit dem λ-Kalkül und von Kurt Gödel, Herbrand, Rózsa Péter und Stephen Kleene ([Kl36]) mit der Theorie berechenbarer Funktionen gelegt wurden.

3.1 Deterministische versus nichtdeterministische Verarbeitung des Empfängers

Bei der Betrachtung interagierender berechenbarer Systeme ist allerdings zusätzlich zu der Frage, wie berechenbare Funktionen grundsätzlich aufgebaut sind, die Frage relevant, welches System welche Informationen zu welcher Zeit zur Verfügung hat und ob die in einem System A ablaufende Berechnung auch aus Sicht des Interaktionspartners B eine solche ist. Mit anderen Worten, nur weil ein System nichts anderes als Berechnungen im Sinne deterministischer Zustandsänderungen ausführt, mithin vollständig durch seinen inneren Zustand und seine Systemfunktion beschrieben werden kann, heißt das nicht, dass es sich aus der Sicht seiner Interaktionspartner als solches darstellt. Wenn die Eingabe und Ausgabe-Partner ständig wechseln, wie es in Kooperationsnetzwerken die Regel ist, dann werden die Interaktionen aus Sicht der Interaktionsteilnehmer im Wesentlichen nichtdeterministischen Charakter haben ([Re12]). Ein nichtdeterministisches Verhalten des Empfängers von Informationen meint, dass der vom Empfänger vorgenommene Zustandsübergang zwar durch die empfangenen Informationen veranlasst, aber keineswegs vollständig durch sie bestimmt wurde. Dies lässt sich von den Informationen, die in einem Kommunikationsschritt ausgetauscht werden, auf alle Informationen, die dem Sender bekannt sind, verallgemeinern.

Auch können dann selbst deterministische Systeme in solchen Interaktionsnetzwerken scheinbar spontanes Verhalten zeigen, was allein mit dem Funktionsbegriff im Sinne einer Abbildung grundsätzlich nicht möglich ist, weil der Abbildung im Falle spontanen Handelns die "Eingabe" fehlt.

Die erste Dimension, entlang derer die Interaktion zwischen Systemen klassifiziert werden kann, bezieht sich daher auf das Verhalten des Empfängers von Informationen aus Sicht des Senders: Ein Empfänger verhält sich entweder deterministisch oder nichtdeterministisch.

3.2 Synchrone versus asynchrone Verarbeitung des Senders

Bei vernetzten Interaktionen kommen die Eingabedaten in der Regel der Reihe nach von verschiedenen Sendern, mit der Folge, dass die interagierenden Systeme keine starren hierarchischen Systemverbünde eingehen.

Die Theorie der berechenbaren Funktionen beschreibt wie eine übergeordnete Funktion aus untergeordneten (wiederum berechenbaren) Funktionen geformt werden kann. Entsprechend lässt sich jeder Berechnung ein Zeitschritt zuordnen, der für die Berechnung

”abgewartet” werden muss. Die Berechnung der übergeordneten Funktion setzt voraus, dass die Berechnung der untergeordneten Funktionen abgeschlossen wurde. Diese Eigenheit ist die inhaltliche Grundlage der Bildung hierarchischer Systemverbünde. Geht Interaktion nun ohne die Bildung von übergeordneter Funktionalität einher, entfällt auch die Notwendigkeit auf Berechnungsschritte anderer Systeme warten zu müssen.

Die zweite Dimension, entlang derer die Interaktion zwischen Systemen klassifiziert werden kann, bezieht sich daher auf das Verhalten des Senders von Informationen aus Sicht des Empfänger: Ein Sender wartet entweder auf das Ergebnis der Berechung des Empfängers (synchrones Verhalten) oder eben nicht (asynchrones Verhalten).

3.3 Zustandslose versus zustandsbehaftete Verarbeitung des Empfängers

Ein Empfänger kann seine Verarbeitung der erhaltenen Informationen (abgesehen von eigenen Informationen) alleine auf diese stützen oder zusätzlich auf in früheren Interaktionen erhaltene Informationen. Den ersten Fall nennt man zustandslose Interaktion, den zweiten Fall zustandsbehaftet.

Dabei ist zwischen dem einzelnen Kommunikationsvorgang und der Interaktion zu unterscheiden. So ist etwa eine Telefon- oder eine HTTP-Verbindung zustandslos. Beide eignen sich aber für die Abwicklung zustandsbehafteter Interaktionen wie Bestellungen, etc., bei der sich jeder Interaktionsschritt auf die vorangegangenen beziehen kann.

4 Acht Kategorien der Interaktion

Die 8 Fälle, die aus der Kombination der drei Dimension entstehen, sind in Tabelle 1 zusammengefasst. Die wesentliche Behauptung ist nun, dass für jeden dieser 8 Fälle bestimmte syntaktische Mittel am besten geeignet sind, um die jeweils zugrunde liegende Interaktionssemantik zu repräsentieren.

Der wichtigste Fall einer **synchronen, deterministischen** Interaktion in der Informatik ist der Aufruf von Funktionen[4]. Die aufgerufene Funktion liefert ihren Wert im Verarbeitungskontext ab, der solange wartet, bis ihre Berechnung fertig ist. Die abgelieferten Informationen erhalten durch die deterministische Interaktion einen imperativen Charakter. D.h. sie beziehen sich auf das was gleich in der Zukunft ganz sicher passieren wird. Entsprechend werden in der Regel für die Beschreibung dieser Interaktionsform Verben in der Befehlsform verwendet. Besitzt der Empfänger der ursprünglich abgelieferten Information keinen eigenen Zustand, so ist es ein einfacher Funktionsaufruf, wie etwa die Berechung eines Sinus: y = sin(x). Für dem Umgang mit zustandsbehafteter Funktionalität hat sich die Objektorientierung herausgebildet, in der die sogenannten Objekte den Zustand kapseln. Etwa bei einem Iterator, der mit einem Startwert initialisiert wird und dessen getNext()-Methode jedesmal den nächsten Iterationswert liefert: iterator.getNext().

[4] Synonyme sind ”Methoden”, ”Operationen” oder ”Prozeduren”

Sender	Empfänger		Syntakt. Mittel zur Beschreibung	Beispiel
	Det.	Zust.		
s	d	zl	Funktionsaufruf	y = sin(x)
s	d	zb	Aufruf einer Objektmethode	y = iterator.getNext()
s	n	zl	Funktionsaufruf mit Ausnahmen die Zufälle abfangen	createConnection(String address) throw TimeoutException
s	n	zb	Aufruf einer Funktion/ Objektmethode mit Ausnahmen	writeTextToFile(File file, String text) throw NoSpaceOnDevice
a	d	zl	Funktionaufruf ohne Rückgabewert/Protokoll	Steuerung eines stabilen Systems, etwa Ausrichtung einer Videokamera oder einer Heizung.
a	d	zb	-	Problematisch, da Asynchronizität des Senders und Zustandsbehaftung des Verhaltens des Empfängers zu Inkonsistenzen führen.
a	n	zl	-	Problematisch, da ohne Zustandsbehaftung des Empfängers sich für die Interaktion kein prozessualer Fortschritt ergibt.
a	n	zb	Protokoll	Netzwerkartige Interaktionen, in denen die beteiligten Systeme wenigstens Eingaben aus zwei verschiedenen Interaktionen miteinander zu koordinieren haben.

Tab. 1: Klassifikation von Interaktionen anhand der Sender- und Empfängercharakteristiken. Ein Sender kann synchron (s) oder asynchron (a) arbeiten, ein Empfänger entweder deterministisch (d) oder nichtdeterministisch (n), sowie zustandslos (zl) oder zustandsbehaftet (zb).

Ein wichtiger Fall **synchroner, nichtdeterministischer** Interaktionen ist der Funktionsaufruf mit Ausnahmen. Dabei wird die im nichtdeterministischen Fall multiple Zuordnung der Eingabe- zu den Ausgabewerten in eine ”gewünschte” und die restlichen ”unerwünschten” Zuordnungen - die ”Ausnahmen” - partitioniert. Ein Beispiel ist etwa das Schreiben eines Textes in ein File. Dies geht in aller Regel gut, es sei denn, die Platte ist voll. D.h. de fakto ist das Verhalten des Filesystems nichtdeterministisch bezogen auf den Befehl, einen Text einem File hinzuzufügen: Es kann funktionieren oder auch nicht. Aber die Funktion ist die Regel und die Störung die Ausnahme. Ich nenne dieses Verhalten quasideterministisch. Das Schreiben eines Files ist zustandsbehaftet, weil es mit der Zeit dazu führt, dass die Platte tatsächlich voll sein wird. Der innere Zustand des Befehlsempfängers kann also dazu führen, dass sich sein Verhalten irgendwann ändert. Im nichtzustandsbehafteten Fall kann kein innerer Zustand den Nichtdeterminismus beeinflussen und entsprechend ist die Ausnahme aus Sicht des Aufrufers Zufall. Ein Beispiel ist etwa die Herstellung einer Kommunikationsverbindung, die an vielen externen Faktoren scheitern kann.

Ein typisches Beispiel von **asynchroner, deterministischer** Interaktion ist die Steuerung (”open loop control”) eines stabilen Systems, etwa einer Heizung. Eine Heizung, beste-

hend aus Brenner, Temperaturfühler, Heizkörper und Regler, lässt sich über ein Temperatursignal steuern. Dabei braucht es eine gewisse Zeit, bis die gewünschte Soll-Temperatur erreicht wird. Um eine neue Soll-Temperatur zu setzen, muss man nicht warten, bis die alte erreicht wurde. Wichtig dabei ist, dass jeder Befehl über kurz oder lang zu einem vom vorhergehenden Zustand unabhängigen Effekt im Befehlsempfänger führt, er also idempotent ist. Wäre dies nicht so, würde die Asynchronizität des Senders und die Zustandsbehaftung des Empfängers zu Inkonsistenzen führen. Zur Beschreibung dieser Form der Interaktion sind einerseits Funktionsaufrufe ohne Rückgabewert geeignet, als auch Dokumentenorientierte Protokolle.

Asynchrone, nichtdeterministische Interaktionen sind die bestimmenden Interaktionsformen in netzwerkartigen Interaktionen. Dort ist die Hauptaufgabe von (deterministischen) Sytemen die Koordination aller Interaktionen, in die sie eingebunden sind [Re12]. Dabei darf in der Regel die Interaktion nicht die Aktion der Systeme bestimmen. Stattdessen ist die lose Beziehung zwischen Interaktion und Aktion Voraussetzung für die Skalierung von Interaktionsnetzwerken, weswegen hier zu Recht von "loser Kopplung" der Systeme gesprochen werden kann.

Diese Interaktionsform ist grundsätzlich zustandsbehaftet. Die abgelieferten Informationen beziehen sich zunächst auf die Vergangenheit, als sie einen Zustandsübergang des sendenden Systems dokumentieren. In einer Interaktion, die gewisse Anforderungen der Konsistenz erfüllt, erfolgt diese Dokumentation in einer Weise die "verständlich" für den Empfänger ist, d.h. dass der Empfänger sie vorbehaltlich seiner eigenen Entscheidungen passend verarbeiten, also einen passenden eigenen Zustandsübergang durchführen kann. Ohne Zustandsbehaftung des Empfängers ergäbe sich für die Interaktion kein prozessualer Fortschritt.

Die Beschreibung dieser Interaktionsform geschieht durch Protokolle[5] (z.B. [Ho91]). Man kann auch nichtdeterministische Zustandsübergänge mit Funktionen beschreiben. Da sich in dieser Interaktionsform der Nichtdeterminismus allerdings nicht absondern lässt, erhält man dann als Ergebnis keine einzelnen Zustandswerte, sondern Mengen von Zustandswerten. Damit wird die Beschreibung von zyklischen Interaktionsketten recht schnell sehr komplex. Dieses Problem wird mit dem Protokoll-Ansatz vermieden.

Die ausgetauschten Informationen haben, wie gesagt, Dokumenten-Charakter. Tatsächlich werden sie von Sender und Empfänger häufig völlig unterschiedlich verarbeitet, besitzen also für beide verschiedene Bedeutungen. So ist etwa ein Preis in einer Geschäftstransaktion für den Käufer die Menge Geld, die er bezahlen muss und für den Verkäufer ist es die Menge Geld, die er erhält. Die ausgetauschten Informationen werden in der Regel mit Hauptwörtern bezeichnet. Typische Beispiele finden sich etwa in Geschäftsinteraktionen: Kaufauftrag, Rechnung, Lieferavis, etc.

[5] Synonyme für "Protokoll" sind "Choreografie" oder auch "Kollaboration" (z.B. [OM13a]).

4.1 Kombination mit der ”anderen” Richtung

Interaktion ist in der Regel nicht einseitig, stattdessen wechseln die Systeme oft die Rollen von Sender und Empfänger. Die Klassen lassen sich symmetrisch oder asymmetrisch kombinieren:

Die symmetrische Kombination (quasi-)deterministischer Interaktionsklassen führt zur rekursiven Supersystembildung [Re10]. Die entstehende Systemfunktion erstreckt sich über die gesamte deterministische Kette. Diese enge Beziehung nenne ich **”Symbiose”**. Sie sollte für die Interaktion von verschiedenen Komponenten möglichst vermieden werden.

Die asymmetrische Kombination aus einerseits synchroner und (quasi-)deterministischer und andererseits asynchroner, nichtdeterministischer, zustandsbehafteter Klasse nenne ich **vertikale Interaktion** oder **”Verwendung”**. Sie beinhaltet zum einen, dass ein per Funktionsaufruf aktivierter Kontext sein Berechnungsergebnis trivialerweise zurücksenden darf. Und es bedeutet weiterhin, dass er sich spontan melden kann, ohne allerdings den Kontext zu kennen, an den er die Indikation seiner Zustandsänderung abliefert, d.h. er kann einen generischen, nicht-Empfänger-spezifischen Event absetzen, etwa ”Tank ist voll”. Dies ist insbesondere bei von mehreren Verwendern benutzen Systemen wichtig, da ein Verwender den Zustand des verwendeten Systems derart verändert haben kann, dass dies für alle potentiellen Verwender interessant sein könnte.

Die symmetrische Kombination der asynchronen, nichtdeterministischen, zustandsbehafteten Interaktionsklasse nenne ich **horizontale Interaktion**. Hier senden sich die Systeme gegenseitig Dokumente im Sinne Empfänger-spezifischer Events. Der ”lose Charakter” der wechselseitigen semantischen Beziehung spiegelt sich dann in der Tatsache wider, dass für die wechselseitige Interaktionssemantik nur die Projektion der Verarbeitung auf die Interaktion relevant ist [Re12].

5 Bemerkungen zur SOA und dem REST

Die beiden von den Proponenten sogenannten Architekturstile der ”Service orientierten Architektur” (SOA) [OA06] und des ”Representational State Transfer” (REST) [Fi00, RR07] haben recht weite Verbreitung gefunden, weswegen ich sie hier in Bezug auf die eingeführte Klassifikation diskutiere.

5.1 Service orientierte Architektur

Die Idee der SOA geht auf R.W. Schulte und Y. V. Natis der Gartner Group zurück. [SN96]. OASIS definiert in ihrem Referenz Modell für die Service orientierte Architektur, Version 1.0 [OA06], eine SOA recht allgemein als ”paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains.”. Ein ”Service” wird definiert als “The performance of work (a function) by one for another.” und als ”mechanism by which needs and capabilities are brought together”.

Tatsächlich wurde die SOA zwischenzeitlich von vielen großen Softwareherstellern im Umfeld der Unternehmenssoftware propagiert, allerdings ohne den erhofften Erfolg speziell bei der Vernetzung von Unternehmensapplikationen.

Die WSDL 1.1 Spezifikation [WSD01] definierte noch, gestützt auf das Muster der gesendeten Nachrichten, vier ”transmission primitives”, die ein sogenannter ”endpoint” unterstützen soll: ”One-way” (der ”endpoint” empfängt eine Nachricht), ”Request-response” (der ”endpoint” empfängt eine Nachricht und sendet eine ”korrelierte” Nachricht zurück), ”Solicit-response” (der ”endpoint” sendet eine Nachricht und erhält eine ”korrelierte” Nachricht zurück) und ”Notification” (der ”endpoint” sendet eine Nachricht). WSDL bezeichnet diese ”transmission primitives” als ”operations”. Damit wird die Semantik dieser ”operations” nicht über ihre Abbildungseigenschaften definiert.

WSDL 2.0 [WSD07a] spricht in Abschnitt 2.2.1 von einer ”Interface component” als einer ”sequences of messages that a service sends and/or receives”. Nach WSDL 2.0 ist eine ”operation” eine ”interaction with the service consisting of a set of (ordinary and fault) messages exchanged between the service and the other parties involved in the interaction”. Auch hier wird die Semantik einer ”operation” nicht über ihre Abbildungseigenschaften definiert.

Die BPMN-Version 2.0.2 [OM13b] bezieht sich formal auf die WSDL 2.0 Spezifikation (wenn auch nur als nichtnormative Referenz) und so ist es keine Überraschung, dass in Abschnitt 8.5.3 nur festgestellt wird: ”An Operation defines Messages that are consumed and, optionally, produced when the Operation is called.” - Wieder keine Definition der Semantik einer ”operation” über ihre Abbildungseigenschaft.

Es lässt sich feststellen, dass die in diesem Artikel vertretene Anschauung von Semantik in der SOA nicht zur Anwendung gekommen ist, ohne dass ein klares alternatives Modell erkennbar wäre.

Die Semantik eines SOA-”Service” könnte ein Funktionsaufruf im Sinne einer synchronen (quasi-) deterministischen Interaktion sein, dessen Semantik dann nach Dana Scott und Christopher Strachey denotational oder entlang der Theorie von Gordon Plotkin operational festgelegt wäre. Wegen des funktionalen Charakters lassen sich dann die Sende- und Empfangsfunktionalität und damit die Erstellung und die Verarbeitung der serialisierten Daten hinter dem Interface verstecken (mit Ausnahme der durch den Transport zusätzlich möglichen Ausnahmen).

Die Semantik eines SOA-”Service” könnte aber auch der Versand eines Dokuments im Sinne einer asynchronen, nichtdeterministischen Interaktion sein, bei der die einzige vom Sender kontrollierte Funktion des Empfängers die des Empfangens ist, die für alle Dokumente aber gleich sein kann. Damit wäre die Semantik aller vom Sender kontrollierten ”SendDataXYZ()” Operationen nach der in diesem Artikel vertetenen Auffassung identisch. Die Semantik der Dokumente ergibt sich hingegen aus den unter ihrer Beteiligung erfolgenden Zustandsübergänge des Empfängers. Da diese nicht Teil der Empfangsoperationen sind, wird die Semantik der Dokumente auch nicht von den Empfangsoperationen berührt.

Die semantische Vagheit trug wohl mit zu Aussagen bei wie ”Everything is a service” (etwa [GI00]) bei und könnte einer der wichtigsten Gründe sein, warum die Konzepte im Umfeld der SOA insgesamt recht komplex geworden sind, häufigen Änderungen unterworfen waren und trotz der Marktmacht der beteiligten Unternehmen nicht den erhofften Durchbruch in der Kommunikation zwischen Unternehmensanwendungen brachten.

5.2 Representational State Transfer

Der sogenannte ”Representational State Transfer” (REST) wird mittlerweile als Alternative zu SOAP und WSDL positioniert, um die Interaktion verteilter Anwendungen im Internet zu beschreiben. Er geht auf Arbeiten von Roy Fielding ([Fi00]) zurück.

Roy Fielding hatte richtig erkannt ([Fi00], S.80), dass das Design Prinzip des Internets das der zustandslosen Dokumentenübertragung war, und versucht, dieses Prinzip auf die Interaktion von vernetzten Applikationen zu übertragen. Entsprechend bezieht sich der REST nicht unmittelbar auf die Operationen eventuell relevanter Anwendungsobjekte (die als Ressourcen bezeichnet werden), sondern allein auf die Manipulation der zu übertragenden Daten mittels der dem HTTP-Protokoll [Fi99] entlehnten Operationen.

Ein Rest-Dienst soll den folgenden 3 Prinzipien genügen (z.B. [RR07]):

- Adressierbarkeit: Jede Ressource wird durch einen eindeutigen URI benamt.
- Zustandslose Kommunikation: Eine REST-Nachricht soll alle Informationen enthalten, die für den jeweilig anschließenden Verarbeitungsschritt des Empfängers notwendig ist.
- Idempotenz: die aufgerufenen Methoden sollen bei jedem Aufruf denselben Effekt haben, so wie es die HTTP-Spzifikation ([Fi99], Abschnitt 9.1.2) für die Operationen GET, HEAD, PUT und DELETE fordert.

Auch bezüglich des REST-Ansatzes lässt sich sagen, dass die Semantik eines REST-Services bezüglich der Anwendungsschicht nicht definiert ist, in diesem Fall allerdings absichtlich. Genau genommen definiert REST nur eine Briefkastensystematik. Im Gegensatz zum traditionellen Briefkasten, der, bis auf das Abliefern der Post, in der Hoheit des Empfängers steht, wird nun diese Hoheit teilweise an den Sender übergeben.

Nimmt man die REST-Prinzipien ernst, dann hat dies insbesondere für netzwerkartig interagierende Applikationen recht merkwürdige Forderungen zur Folge: Die Ablieferung eines Kaufauftrags muss jederzeit wiederholt (idempotentes PUT) und jederzeit widerrufen werden können (idempotentes DELETE). Und anstelle des bewährten einen Briefkasten für alle abgesendeten Dokumente erhält man eine Vielzahl von verschiedenen Briefkästen, für jede sogenannte ”Ressource” einen eigenen. Auch dürfte etwa eine Überweisung keine Referenz auf den Kaufauftrag oder eine (Teil-) Lieferung enthalten, da die Interaktionen ja zustandslos sein sollen.

Damit ist REST in gewissem Sinn ein Zwitter, der nur Teile einer Objekt-Zugriffsschicht in die Administrationsschicht übernommen hat. Entsprechend gehen wichtige Eigenschaften für die Beschreibung netzwerkartiger Interaktionen verloren und für die Beschreibung von Funktionsaufrufen sind weiterhin wesentliche Ergänzungen notwendig, wie das OData-Protokoll zeigt [ODa13].

6 Abschließende Bemerkungen

Die vorgestellte Klassifikation entlang der Sender-bezogenen Synchronizität und der Empfänger-bezogenen Determiniertheit und Zustandsbehaftung ist vollständig. D.h. jede wohldefinierte Interaktion lässt sich entsprechend zuordnen.

Die in der Informatik sehr verbreitete Vorstellung, dass die Interaktion von Systemen allein durch benamte Operationen zu beschreiben ist, ist damit unzureichend. Sie liegt bemerkenswerter Weise etwa auch den Prozesskalkülen von Charles H. Hoare [Ho04] als auch von Robin Milner [Mi89, MPW92] zugrunde. Tatsächlich impliziert die in diesem Artikel vertretene Auffassung von Semantik unmittelbar, dass die Angabe von Operationen bzw. sogar ihre feste Verküpfung mit Daten im Sinne der Objektorientierung eine sehr starke semantische Festlegung darstellt. Diese Festlegung ist in netzwerkartigen Interaktionen ungünstig, da ihre Flexibilität ganz wesentlich auf der Interpretationshoheit der Informationsempfänger beruhen.

Ein wesentliches Ergebnis dieses Artikels ist ein klares und einfaches Kriterium, um horizontale und vertikale Interaktionen eindeutig zu unterscheiden. Insbesondere ist das Absetzen generischer Events kompatibel zur Steuerung zustandsbehaftete Objekte mit traditionellem funktionalen APIs. Hier wäre denkbar, dass Komponentenmodelle das State-Pattern [Ga95] deklarativ unterstützen und so deutlich machen, welche Verhaltensänderung gegenüber den Objekt-Verwendern per Event generisch signalisiert werden kann.

Es folgt unmittelbar, dass Komponentenmodelle, die keine syntaktischen Mittel zur Verfügung stellen, um Protokolle zu beschreiben, für Komponenten, die in Interaktionsnetzwerken Verwendung finden sollen, ebenfalls unzureichend sind. Ivica Crnkovic, Severine Sentilles, Aneta Vulgarakis und Michel R. V. Chaudron [Cr11] unterscheiden in ihrem Überblick über Komponentenmodelle zwischen ”Operation based” und ”port based” Interface-Unterstützung und zeigen damit auf, dass viele wichtige Komponentenmodelle bisher tatsächlich keine Protokolldeklarationen unterstützen.

Komponentenmodelle, die nur prozedurale oder objektorientierte Mittel zur Verfügung stellen, eignen sich daher nur zur Beschreibung von zusammengesetzten Systemen, die in einer hierarchischen Beziehung stehen. Und ohne die Möglichkeit der Beschreibung generischer Eventkonsumption der Verwender auch dazu nur eingeschränkt.

Bei zyklischen Abhängigkeiten von Komponenten, die durch funktionale Schnittstellen beschrieben werden, entstehen rekursive Funktionen über Komponenten hinweg, die das Design der Komponenten sehr stark wechselseitig - symbiotisch - abhängig machen. Das heißt nicht, dass sich mit prozeduralen oder objektorientierten Mitteln keine nichtdetermi-

nistischen Beziehungen zwischen Komponenten aufbauen ließen, sondern nur, dass man diese Beziehungen mit den syntaktischen Mitteln des Komponentenmodells dann nicht adäquat beschreiben kann. Ein gutes Beispiel ist etwa die sogenannte AJAX-Architektur für Web-Applikationen, bei der über die Beziehung zwischen dem Web-Client und dem Server nur soviel deklarativ bekannt ist, als sie über HTTP kommunizieren. Diese Unterspezifikation des Verhältnisses der beiden Systeme stellt in einer Umgebung, in der etwa die Laufzeitumgebung des Web-Client als nicht-vertrauenswürdig eingestuft werden muss, ein enormes Sicherheitsrisiko dar.

Ohne eine deklarative Unterscheidung horizontaler und vertikaler Interaktion ist auch jede formale Analyse zyklischer Abhängigkeiten auf Komponentenebene nur begrenzt aussagefähig.

Dieses Defizit verbreiteter Komponentenmodelle ist angesichts der Tatsache, dass ab einer bestimmten Komplexität der Komponenten diese in der Regel in nichtdeterministische Interaktionen eingebunden sind, erstaunlich. Protokolle ermöglichen ganz wesentlich "Programming in the Large", um die Terminologie von Frank DeRemer and Hans Kron aufzugreifen [DK75, Si04].

Danksagung: Mein besonderer Dank geht an Christoph Rechsteiner. Die vorgestellte Klassifizierung entstand in einer ersten Form in einer gemeinsamen Diskussion zwischen uns beiden im Herbst 2014 im Zusammenhang unserer Arbeiten im Umfeld der Standardisierung zu Industrie 4.0.

Literatur

[Au62] Austin, John Langshaw: How to Do Things with Words. Cambridge (Mass.), 1962.

[BM12] BMBF, Referat Grundsatzfragen der Innovationspolitik: . Bericht der Bundesregierung, Zukunftsprojekte der Hightech-Strategie (HTS-Aktionsplan), 2012.

[Ch36] Church, Alonzo: An unsolvable problem of elementary number theory. American Journal of Mathematics, 58:345 – 363, 1936.

[Cr11] Crnkovic, Ivica; Sentilles, Séverine; Vulgarakis, Aneta; Chaudron, Michel R. V.: A classification framework for software component models. Software Engineering, IEEE Transactions on, 37(5):593–615, 2011.

[DK75] DeRemer, Frank; Kron, Hans: Programming-in-the Large Versus Programming-in-the-small. SIGPLAN Not., 10(6):114–121, April 1975.

[Fi99] Fielding, R.; Gettys, J.; Mogul, J.; Frystyk, H.; Masinter, L.; Leach, P.; Berners-Lee, T.: , RFC 2616, Hypertext Transfer Protocol – HTTP/1.1. `http://tools.ietf.org/html/rfc2616`, 1999. Aufruf 2015-01-17, überholt seit 2014 durch RFCs 7230, 7231, 7232, 7233, 7234, and 7235.

[Fi00] Fielding, Roy: Architectural Styles and the Design of Network-based Software Architectures. Dissertation, University of California, Irvine, 2000.

[Fr92] Frege, Gotlob: Über Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik, 100:25–50, 1892.

[Ga95] Gamma, Erich; Helm, Richard; Johnson, Ralph; Vlissides, John: Design Patterns, Elements of Reusable Object-Oriented Software. Addison-Wesley, 1. Auflage, 1995.

[GI00] Gottschalk, Karl; IBM Web Services Architecture Team: , Web Services architecture overview. http://www.ibm.com/developerworks/webservices/library/w-ovr/, 2000. Augerufen am 2015-01-17.

[Ho91] Holzmann, Gerard J.: Design and validation of computer protocols. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1991.

[Ho04] Hoare, Charles.A.R.: Communicating Sequential Processes. Prentice Hall, 1985/2004.

[I4015] Umsetzungsstrategie Industrie 4.0, Ergebnisbericht der Plattform Industrie 4.0. http://www.plattform-i40.de/sites/default/files/150410_Umsetzungsstrategie.pdf, 2015.

[ITU94] Information Technology - Open Systems Interconnection – Basic Reference Model. http://handle.itu.int/11.1002/1000/2820, 1994. Früher "CCITT Recommendation", Identischer Standard: ISO/IEC 7498-1:1994 (Common).

[Kl36] Kleene, Stephen: General recursive functions of natural numbers. Mathematische Annalen, 112(5):727–742, 1936.

[KLW11] Kagermann, Henning; Lukas, Wolf-Dieter; Wahlster, Wolfgang: Industrie 4.0: Mit dem Internet der Dinge auf dem Weg zur 4. industriellen Revolution. VDI Nachrichten, 13, April 2011.

[LW94] Liskov, Barbara H.; Wing, Jeannette M.: A behavioral notion of subtyping. ACM Trans. Program. Lang. Syst., 16(6):1811–1841, 1994.

[LW01] Liskov, Barbara H.; Wing, Jeannette M.: Behavioural subtyping using invariants and constraints. In (Bowman, Howard; Derrick, John, Hrsg.): Formal methods for distributed processing: a survey of object-oriented approaches, S. 254–280. Cambridge University Press, New York, NY, USA, 2001.

[Mi89] Milner, Robin: Communication and Concurrency. Prentice Hall, 1989.

[MPW92] Milner, R.; Parrow, J.; Walker, D.: A calculus of mobile processes (parts I and II). Information and Computation, 100(1):1–77, 1992.

[OA06] OASIS: , Reference Model for Service Oriented Architecture 1.0. http://docs.oasis-open.org/soa-rm/v1.0/, 2006. Aufgerufen am 2015-01-17.

[ODa13] Open Data Protokoll, Version 4.0. http://www.odata.org/documentation/odata-version-4-0/, 2013. Aufgerufen am 2015-01-17.

[OM13a] OMG: , Business Process Model and Notation. http://www.omg.org/spec/BPMN/2.0.2, 2013. Accessed 2015-01-17, also published as ISO/IEC 19510:2013.

[OM13b] OMG: , Business Process Model and Notation. http://www.omg.org/spec/BPMN/2.0.2, 2013. Aufgerufen am 2015-01-17, auch veröffentlicht unter ISO/IEC 19510:2013.

[Pl81] Plotkin, Gerald D.: A structural approach to operational semantics. Bericht, Aarhus University Computer Science Department, 1981. DAIMI FN-19.

[Pu75] Putnam, Hilary: The meaning of meaning. In (Gunderson, Keith, Hrsg.): Language, Mind, and Knowledge. University of Minesota Press, Mineapolis, USA, 1975. German translation: "Die Bedeutung von Bedeutung", 2. Ed. printed by Klostermann Verlag 1990, Frankfurt a. Main, Deutschland.

[Re08] Reich, Johannes: Modelling the Interaction of Distributed Systems as Protocols. In: MCETECH. IEEE Computer Society, Los Alamitos, CA, USA, S. 16–24, 2008.

[Re10] Reich, Johannes: Finite System Composition and Interaction. In (Fähnrich, Klaus-Peter; Franczyk, Bogdan, Hrsg.): 40. GI Jahrestagung (2). Jgg. 176 in LNI. GI, S. 603, 2010.

[Re12] Reich, Johannes: Processes, roles and their interactions. In (Reich, Johannes; Finkbeiner, Bernd, Hrsg.): Proceedings Second International Workshop on Interactions, Games and Protocols, IWIGP 2012, Tallinn, Estonia, 25th March 2012. Jgg. 78 in EPTCS, S. 24–38, 2012.

[RNI01] RosettaNet Implementation Framework: Core Specification, V02.00.00, 2001. http://xml.coverpages.org/RNIF-Spec020000.pdf, aufgerufen am 2007-3-11.

[RR07] Richardson, Leonard; Ruby, Sam: RESTful Web Services, Web services for the real world. O'Reilly Media, 2007.

[Se69] Searle, John: Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press, 1969.

[Sh48] Shannon, Claude E.: A Mathematical Theory of Information. Bell System Technical Journal, 27:379–423, 623–656, 1948.

[Si04] Singh, Munindar P; Chopra, Amit K; Desai, Nirmit; Mallya, Ashok U: Protocols for processes: programming in the large for open systems. ACM Sigplan Notices, 39(12):73–83, 2004.

[SM09] Savolainen, Juha; Myllärniemi, Varvana: Layered architecture revisited - Comparison of research and practice. In: Joint Working IEEE/IFIP Conference on Software Architecture 2009 and European Conference on Software Architecture 2009, WICSA/ECSA 2009, Cambridge, UK, 14-17 September 2009. IEEE, S. 317–320, 2009.

[SN96] Schulte, Roy W.; Natis, Yefim V.: "Service-Oriented" Architectures, Part 1 and 2. SSA Research Notes SPA-401-068, -069, Gartner Group, 1996.

[SS71] Scott, Dana; Strachey, Christopher: Toward a mathematical semantics for computer languages. In (Fox, J., Hrsg.): Proc. Symp. Computers and Automata. Poytechnic Inst. of Brooklyn Press, 1971. Also Technical Monograph PRG-6, Programming Research Group, Oxford University.

[Tu36] Turing, Alan M.: On Computable Numbers, With an Application to the Entscheidungsproblem. Proceedings of the London Mathematical Society, Series 2, 42:230–265, 1936.

[UMM03] UN/CEFACT Modelling Methodology (UMM) User Guide, V20030922, 2003.

[WSD01] Web Services Description Language (WSDL) 1.1. `http://www.w3.org/TR/wsdl`, 2001. Aufgerufen am 2015-01-17.

[WSD07a] Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language. `http://www.w3.org/TR/wsdl20/`, 2007. Aufgerufen am 2015-01-17.

[WSD07b] Web Services Description Language (WSDL) Version 2.0: Additional MEPs. `http://www.w3.org/TR/wsdl20-additional-meps`, 2007. Aufgerufen am 2015-04-15.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Enabling Industry 4.0 with holobuilder

Maximilian Speicher[1], Kristina Tenhaft[2], Simon Heinen[3] and Harry Handorf[4]

Abstract: Industry 4.0 denotes the emergence of smart factories with a strong focus on communication between humans, machines and products, which is supported by cyber-physical systems [HPO15]. Among such systems are augmented reality applications, which have great value for the assembly, monitoring and maintenance of machines. Yet, to date there are no convenient means for end-users to create augmented reality content in a simple manner. We introduce *holobuilder*—a novel platform for the creation of interactive instruction manuals in Industry 4.0 settings. Additionally, the created augmented reality scenarios can be enhanced with OPC UA capabilities, which enables communication between devices displaying holobuilder content and the monitored machines. We assess our platform against six design principles for Industry 4.0 systems and show that all of them are successfully met. Moreover, we demonstrate the effectiveness and feasibility of holobuilder in two real-world case studies involving a thermostat as well as a machine handling gears in an assembly line.

Keywords: 3D, Augmented Reality, Cyber-Physical Systems, End-User Design, Industry 4.0, Interactive Instruction Manuals, OPC UA, Smart Factory, Virtual Reality

1 Introduction

Industry 4.0 (German: Industrie 4.0) is a term originally coined by the German federal government in the context of their *High-Tech Strategy 2020 for Germany* [HPO15, Hi15]. It is used to describe the fourth—and currently happening—industrial revolution. At the core of this revolution stands the *smart factory* in which humans, different machines and products communicate with each other [Pl15].[5] Based on a recent and comprehensive literature review, [HPO15] define Industry 4.0 as follows:

> "Industrie 4.0 is a collective term for technologies and concepts of value chain organization. Within the modular structured Smart Factories of Industrie 4.0, [cyber-physical systems] monitor physical processes, create a virtual copy of the physical world and make decentralized decisions. Over the [Internet of Things], [cyber-physical systems] communicate and cooperate with each other and humans in real time. Via the [Internet of Services], both internal and cross-organizational services are offered and utilized by participants of the value chain."

[1] bitstars GmbH, Hanbrucher Straße 40, 52064 Aachen, max@bitstars.com
[2] bitstars GmbH, Hanbrucher Straße 40, 52064 Aachen, kristina@bitstars.com
[3] bitstars GmbH, Hanbrucher Straße 40, 52064 Aachen, simon@bitstars.com
[4] bitstars GmbH, Hanbrucher Straße 40, 52064 Aachen, harry@bitstars.com

[5] Original: "Im Mittelpunkt der Industrie 4.0 steht die intelligente Fabrik, die Smart Factory, in der Menschen, unterschiedliche Maschinen und Produkte miteinander kommunizieren." [Pl15].

As can be seen, Industry 4.0 strongly focuses on more efficient machine-to-machine (M2M) as well as human–machine communication that lead to added value in the process chain. Along with the first specific definition of the term above, [HPO15] describe six *design principles* for Industry 4.0 intended for both, researchers and practitioners.

Among the core components of Industry 4.0 are *cyber-physical systems* (CPS) [HPO15], i.e., systems that result from a "fusion of the physical and the virtual world" [Ka14]. One type of such CPS are Augmented Reality (AR) applications, which are "a live direct or indirect view of a physical, real-world environment whose elements are *augmented* (or supplemented) by computer-generated sensory input such as sound, video, graphics or GPS data."[6] An example application of AR would be the Cologne Cathedral that is being viewed through the camera of a user's smartphone that at the same time automatically displays information about the building. Also in industrial contexts, one can greatly benefit from AR systems, e.g., when it comes to the assembly, monitoring and maintenance of machines in smart factories. Yet, there is still a lack of adequate solutions enabling the simple creation of AR content for end-users without specific training.

We introduce *holobuilder*—a novel web platform that allows end-users to create and share arbitrary AR content in terms of "3D projects". Thus, we enable *end-user design* for AR content in analogy to end-user programming. While holobuilder is intended as a *general* platform, in the context of Industry 4.0 it is particularly suited for the following two use cases:

Interactive Instruction Manuals, which allow users to view a machine through the camera of their mobile device, with context-aware instructions automatically displayed through AR means.

OPC UA, the *Unified Architecture* (UA) developed by the *Object Linking and Embedding for Process Control* (OPC) Foundation. That is, a protocol for M2M communication described as "the interoperability standard for the secure and reliable exchange of data in the industrial automation space and in other industries" [OP15].

Regarding these use cases, our platform has demonstrated its effectiveness and feasibility in two case studies. First, we have implemented a fully working interactive instruction manual for a thermostat using holobuilder alone. Second, we have successfully created an OPC UA–enabled holobuilder project for monitoring a machine handling gears in an assembly line.

In the following, we will give an overview of important background concepts and related work in Section 2 (**Background & Related Work**). Particularly, the six design concepts for Industry 4.0 described by [HPO15] will be introduced in detail. The subsequent section (**holobuilder**) presents holobuilder and gives details about the platform's architecture and implementation. Moreover, it elaborates on particular design choices to improve usability and user experience for end-users. Section 4 (**Evaluation**) evaluates our platform against

[6] `http://www.wikiwand.com/en/Augmented_reality` (accessed 2015/04/23).

the Industry 4.0 design concepts ([HPO15]) to substantiate its feasibility and particular usefulness for Industry 4.0 scenarios. Subsequently, Section 5 (**Case Studies**) in detail describes the two case studies mentioned above before discussing our approach and giving concluding remarks in Section 6 (**Discussion & Conclusions**).

2 Background & Related Work

The following gives an overview of existing work and research that is related to our platform, specifically in the areas of interactive instruction manuals and OPC UA.

2.1 Interactive Instruction Manuals

REFLEKT ONE [RE15] is a general software for creating AR applications in industry, which include interactive instruction manuals. The software allows to re-use existing 2D and 3D content and caters for the process of "linking traditional content and diagnostic data to CAD and 3D animations" [RE15] and "Connect[ing] virtual content to real world objects" [RE15]. In contrast to holobuilder, REFLEKT ONE is not realized in terms of a web application that can be uitilized for creating 3D content with a modern web browser alone. Hence, it is oriented at specific industry solutions rather than (general) end-users. In particular, the software does not offer predefined 3D objects and animations. Rather, it focuses on transferring existing 3D content (e.g., from CAD applications) into an AR setting that can thus be arbitrarily customized.

NGRAIN[7] have cooperated with *Lockheed Martin*[8] for providing an AR-based *computerized maintenance management system* (CMMS) for airplanes [He15a]. That is, Lockheed Martin's engineers are provided with virtual real-time information that allow them to maintain planes in a shorter amount of time with less potential for errors [He15a].[9] Compared to our system, this is a very specific and individual industry solution, particularly because it involves very large 3D models of Lockheed Martin airplanes [He15a]. [He15a] as well states that "real-time updates on factors like temperature, pressure and other indicators are available, including dynamically-changing parts colouring". However, whether this happens through OPC UA does not become evident from the article.

The *Plant@Hand*[10] project of the *Fraunhofer IGD* institute provides a suite of solutions for different tasks in the context of Industry 4.0. Among other things, these include interactive instruction manuals (*Plant@Hand Montage Assistent*), which are context-sensitive and support novel devices such as smart watches. Moreover, the *Plant@Hand3D* subproject aims at optimizing process flows through 3D visualizations of production data. Yet, these solutions are not particularly aimed at or built on AR scenarios.

[7] `http://www.ngrain.com/` (accessed 2015/04/28).

[8] `http://www.lockheedmartin.com/` (accessed 2015/04/28).

[9] "'In short, our customers use our technology to equip their mobile workers with access to the right data at the right time on the right device,' says Director of Product Management Barry Po. 'We provide real-time assistance and allow them to access all the information they need to do their jobs to the fullest extent possible.'" [He15a].

[10] `http://www.igd.fraunhofer.de/node/2387` (accessed 2015/04/28).

ioxp[11] provide interactive AR manuals that display necessary steps in terms of semi-transparent video overlays. That is, the correct execution of the step is recorded beforehand while the system analyzes the state of the workplace before and after the execution. Based on automatic task segmentation they provide a solution that does not require the user to indicate when a step is finished and gives automatic feedback on whether the execution was correct or not [PS12]. Yet, for robustness it is optimal to record the same step multiple times [PS12]. The process of creating ioxp's AR manuals is fundamentally different from holobuilder since it is strongly focused on the involved video captures. Furthermore, they make no statements about the integrability of their solution with OPC UA. OPC UA will be addressed in detail in the following section.

2.2 OPC UA

OPC UA is short for *OPC Unified Architecture*, a standard for machine-to-machine communication developed by the *Object Linking and Embedding for Process Control* (OPC) Foundation[12]. The main goal of the standard is to provide a protocol that is independent of the manufacturers of the communicating machines [Bi14]. In order to support a wide range of available systems, OPC UA is based on three core functionalities: *transport*, *meta model* and *services* [Bi14]. According to [Bi14] "Transport and Services provide the protocol (the 'How') which is used to exchange data, while the Meta Model can be understood as the information (the 'What') that has to be transmitted." Based on the meta model we can describe arbitrary physical systems as OPC UA models that are based on a specified, generic OPC UA object [Bi14]. Such OPC UA models can then be represented by a corresponding OPC UA server [Bi14]. In the context of holobuilder, we have to define such a server in terms of the following standardized scheme to be able to create OPC UA–enhanced AR content:

```
opc.tcp://MyServerPcName:4840
```

The server itself is XML-based and transfers data to the client via TCP [He15b]. It uses a secure, reliable and service-oriented architecture to ultimately deliver machine and process data to the monitoring and controlling systems [He15b].

In [He15b] the use of OPC UA in industry scenarios—particularly in combination with novel wearable devices—has been coined *HMI 4.0*, i.e., Human–Machine Interaction 4.0. This refers to the fact that with the help of OPC UA and wearables, smart factories can reach a new level of interaction between machines and mechanics/engineers. While AR enables a *visual* fusion of the real and the virtual world, the additional use of OPC UA furthermore leads to a fusion on the *information* level. Hence, one core aim of holobuilder is to leverage the advantages of AR-driven interactive instruction manuals *and* OPC UA in Industry 4.0 settings.

[11] `http://www.ioxp.de/` (accessed 2015/04/28).
[12] `https://opcfoundation.org/` (accessed 2015/04/28).

2.3 Design Principles for Industry 4.0

In their working paper, [HPO15] present the six core design principles for Industry 4.0, which are based on a recent and comprehensive literature review of 51 academic publications. These design principles are summarized in the following [HPO15]:

Interoperability It is crucial for all parties involved in Industry 4.0—i.e., people, companies and CPS—to communicate among each other. For this, it is necessary to provide adequate standards, e.g., OPC UA.

Virtualization "CPS are able to monitor physical processes" [HPO15]. For this, one requires appropriate interfaces between the virtual and the real world for creating corresponding mappings.

Decentralization Machines are controlled and act autonomously in a smart factory. The need for centralization is reduced to a minimum, i.e., the case of failures [tHO14].

Real-Time Capability "For organizational tasks it is necessary that data is collected and analyzed in real time." [HPO15]

Service Orientation Services enabling Industry 4.0 are provided internally as well as externally, i.e., they can be used by other participants in the context of Industry 4.0.

Modularity Industry 4.0 is also enabled by modularized systems that can react flexibly to changing contexts, e.g., "in case of seasonal fluctuations or changed product characteristics" [HPO15].

These well-founded principles are intended as a starting point for the implementation of Industry 4.0 scenarios as well as further research [HPO15]. Yet, they also make it possible to assess the suitability of systems for Industry 4.0.

3 holobuilder

holobuilder is intended as a general web-based platform for the creation and consumption of arbitrary AR content by end-users. This means, in analogy to end-user programming, our particular focus is on good user experience and usability for users with no special training in creating AR content. In the context of Industry 4.0, engineers and mechanics responsible for the assembly, maintenance and monitoring of machines in smart factories acquire the position of such end-users, as is implied by the original use case of our platform (Fig. 1). First, in terms of users that *create* AR content, e.g., in the form of an interactive instruction manual. Second, in terms of users that *consume and apply* the AR content created with holobuilder.

Our platform allows end-user to create "3D projects", which can be seen as a 3D analogy to presentations created with, e.g., *Microsoft PowerPoint*; however, with three-dimensional

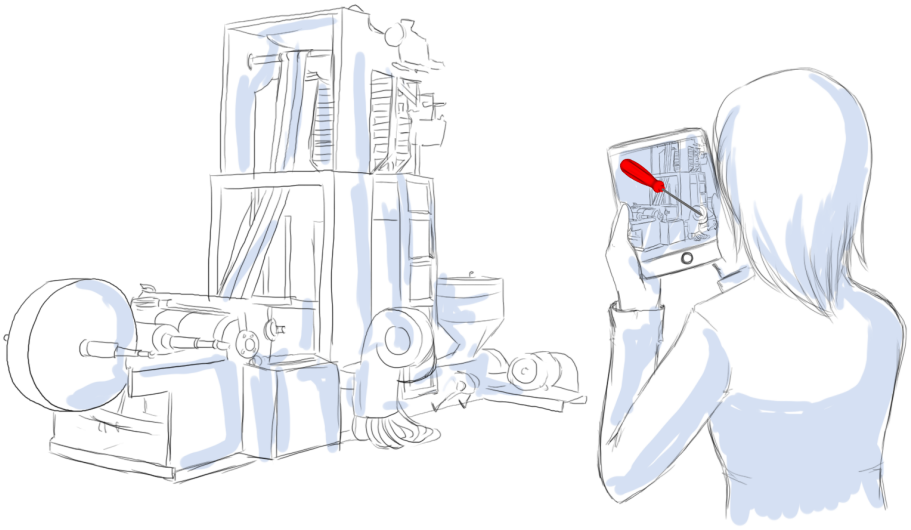


Fig. 1: Early design concept of the original use case of holobuilder: The user views a machine through the camera of his mobile device. The view of the real world is automatically augmented with 3D objects indicating the next step necessary for maintaining the machine.

rooms instead of two-dimensional slides [Sp15]. Yet, if 2D information is required in a project, holobuilder as well supports two types of slides along its 3D rooms, i.e., *text* and *checklist* slides. holobuilder acts as both, the *editor* for creating and the *player* for consuming 3D content and can switch between these two modes at any time.

Interactive Instruction Manuals For creating AR content like interactive instruction manuals, a marker has to be added to an empty room. This marker could be, e.g., a bar code or QR code that is placed on a machine to be maintained (in this case the marker would be a 2D object). Alternatively, holobuilder supports 3D markers, which would be a point cloud of the machine. This point cloud has to be obtained beforehand by scanning the machine using an appropriate 3D scanner. Subsequently, all virtual 3D objects added to the room, e.g., a screwdriver, are placed in relation to the given marker. Since holobuilder recognizes the marker in the real world[13], in this way it is possible to place 3D objects at the exact necessary position in the AR scene. For instance, the exemplary screwdriver can be placed above the screw that needs to be removed. Since they are placed in relation to

[13] Using simultaneous localization and mapping techniques (SLAM; cf. `http://www.wikiwand.com/en/Simultaneous_localization_and_mapping`, accessed 2015/04/28).

the marker, all 3D objects added to the room remain at their assigned global coordinates as the mechanic moves their mobile device around the machine in the real world. Furthermore, the virtual objects can be animated (rotate, bounce, spin etc.) in order to indicate the exact way in which the mechanic should use the tools shown. Also, it is possible to attach click events to the inserted objects, which then show additional information, open a given website URL or send an e-mail, among other things.

Each step in an interactive instruction manual has to be realized as a separate room in the 3D project. The mechanic has to indicate when a step is finished and is subsequently presented with the next room by holobuilder.

OPC UA Similar to the above, it is possible to enhance holobuilder's 3D projects by adding OPC UA capabilities. That is, for a machine to be monitored, one has to set up a corresponding AR scenario including the aforementioned markers. Then, in analogy to the click events that can be attached to 3D objects, one can define OPC UA events if the project is OPC UA–enabled and has a valid server set. For instance, the color of an inserted object can be changed depending on the state of a boolean variable (that is communicated by the monitored machine via the given OPC UA server). This could be a green 3D arrow for a positive state and a red one for a negative state. To give another example, it is possible to add dynamic information, e.g., in terms of a virtual text object displaying the current temperature of the machine.

This kind of information flow as well works into the other direction, which means that variable changes can be triggered by the holobuilder project. For instance, a certain component of the monitored machine could be augmented with a virtual button. Upon clicking that button a boolean variable could be toggled that switches the component on or off. This variable change would be communicated to the machine via the given OPC UA server and then take effect. In this way, holobuilder does not only cater for the monitoring of machines via OPC UA, but also enables remote control capabilities.

3.1 Design Decisions

Since holobuilder is intended as a general tool for *end-users* to create 3D content, our platform is based on a number of core design decisions ensuring an optimal experience for users with no specific training in this domain.

Enhancement of Known Concepts holobuilder builds on familiar concepts to acquaint end-users with creating 3D content. We adopted the core principle of *Microsoft PowerPoint*—creating slides that can have different types and layouts and interconnecting them—that is already well-known to the majority of users[14] and transferred it into the three-dimensional setting [He15b].

[14] `https://infogr.am/PowerPoint-usage-and-Marketshare` (accessed 2015/04/27).

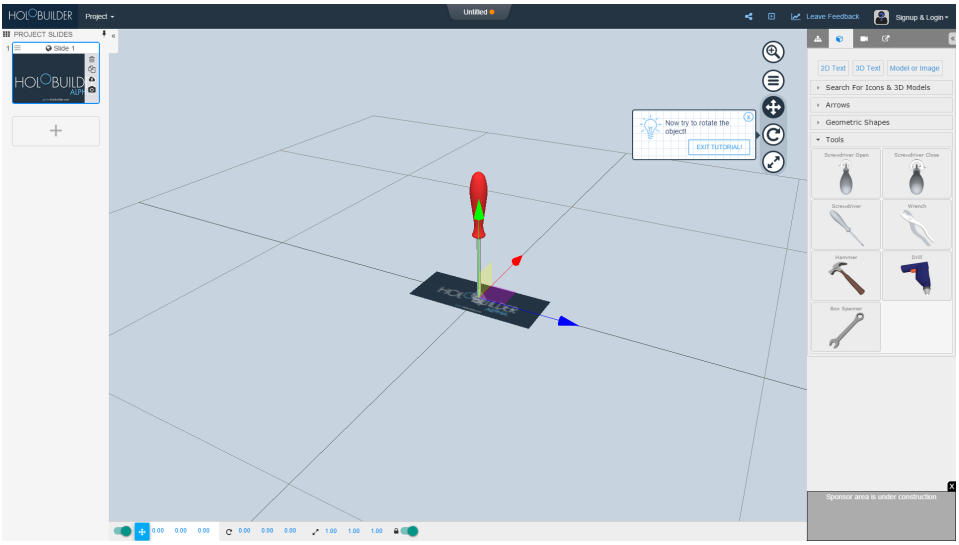


Fig. 2: Screenshot of the holobuilder UI showcasing a tutorial tooltip as well as a selection of the available built-in 3D objects for Industry 4.0 use cases.

First User Experience A step-by-step tutorial guides first-time users through all steps of the process from starting a new project in holobuilder to adding, manipulating and animating virtual 3D models (Fig. 2). The project created in this tutorial serves as a prototype from which more advanced projects can be derived.

Tooltips & Hints Since dealing with 3D content is not a trivial task for end-users, tooltips and hints are present across all stages in the process of creating a project (Fig. 2). Particularly, guidance is necessary when it comes to adding the correct marker type to a room, or when uploading custom 3D objects, as particular file types are required.

Predefined Objects & Animations We provide users with a set of predefined 3D objects and animations that cover major use cases (Fig. 2). Among the provided objects are custom text, arrows, screwdrivers as well as a variety of geometric shapes. As a part of future work, we intend to extend our collection to a "clip art gallery for 3D content" through both, internal and external input. In case a desired object is not provided by holobuilder, users are given the possibility to upload arbitrary 3D models from their computer.

3D Object Search Since not all necessary objects for all use cases can be covered by our gallery (or are present on users' computers), we have included a search for arbitrary 3D objects. This search has been realized using an external API[15] through which we ensure that only objects with the correct file format can be imported into holobuilder.

[15] `https://www.yobi3d.com/#!/` (accessed 2015/04/27).

3.2 Architecture & Implementation

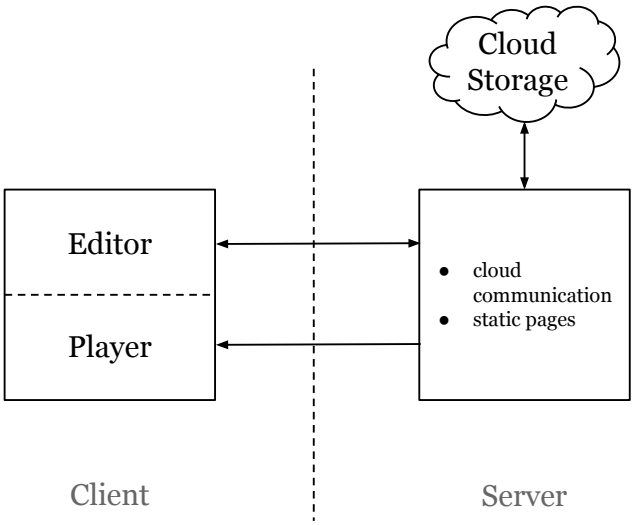


Fig. 3: High-level architecture of the holobuilder platform.

Our platform comprises a number of components on the client as well as the server side. In particular, there are two separate components for creating (*Editor*, Fig. 3) and consuming (*Player*, Fig. 3) 3D projects. The player is commonly used on mobile devices—particularly in AR scenarios—but is not restricted to that class of devices.

holobuilder has been implemented in terms of a *web application* to ensure that simply a modern web browser is required for end-users to create 3D content. The architecture of the platform is highly oriented at the asynchronous paradigm, which means that most of the core logic resides on the client side (Fig. 3). Hence, the implementation of the platform is mainly based on *Backbone.js*[16] in combination with *RequireJS*[17] and *WebGL*[18]. In this way, a slow or unstable Internet connection does not affect the actual use of the holobuilder editor. The server-side components come into play when loading (*Editor*, *Player*) or storing (*Editor*) 3D projects (Fig. 3). All projects created with holobuilder are stored in the cloud to guarantee worldwide availability, but can be alternatively downloaded to the users' computer (in terms of a `.holo` file that can be interpreted by the holobuilder player). Therefore, the holobuilder server is mainly responsible for communication with the cloud provider and serving a number of static pages (such as different landing pages). Its implementation is based on *Java Servlets* in combination with *JSP*.

The live version of holobuilder can be accessed at `https://www.holobuilder.com/`.

[16] `http://backbonejs.org/` (accessed 2015/04/27).
[17] `http://www.requirejs.org/` (accessed 2015/04/27).
[18] `https://www.khronos.org/webgl/` (accessed 2015/04/29).

4 Evaluation

In the following, we assess holobuilder against the six design principles for Industry 4.0 defined by [HPO15] (cf. Sec. 2).

✔ **Interoperability** holobuilder supports interoperability on two levels of abstraction. First, our platform enables machine-to-machine communication through OPC UA. Second, holobuilder itself has been realized as a web app running in any modern web browser. Hence, any appropriate mobile device can be used to monitor machines in a corresponding holobuilder-based AR scenario.

✔ **Virtualization** Since our platform enables the use of AR in smart factories, it clearly fuses the physical and the virtual world (which characterizes CPS according to [Ka14]). Moreover, it leverages the advantages of OPC UA for monitoring physical processes by communicating with the monitored machines.

✔ **Decentralization** holobuilder supports the decentralization of smart factories since, e.g., interactive instruction manuals can be created separately for each machine (optionally with OPC UA capabilities). This facilitates decentralized monitoring and maintenance.

✔ **Real-Time Capability** Naturally, AR scenarios such as interactive instruction manuals are applied in real time since the corresponding machine is viewed live through a camera. Also, in the case of an OPC UA–enabled holobuilder project, machine parameters are communicated in real time.

✔ **Service Orientation** Projects created with holobuilder are not specific to a machine *and* factory. That is, as long as the same marker is being detected by our platform, interactive instruction manuals and OPC UA–enabled projects can be applied across different smart factories. Also, holobuilder itself is a general platform not bound to any particular context. It can be used in any smart factory with any appropriate mobile device.

✔ **Modularity** Our platform supports modular systems in smart factories, as it simplifies the process of adding, changing and monitoring individual components of machines.

As becomes evident from the above, holobuilder has the potential of being successfully applied in Industry 4.0 settings—particularly smart factories. Therefore, we have taken the evaluation of our platform one step further and implemented two case studies that are presented in the following section.

5 Case Studies

To demonstrate the feasibility and effectiveness of holobuilder, we have used it in two case studies that are typical scenarios in Industry 4.0 contexts. First, Kristina Tenhaft—in the

context of her Master's thesis—has implemented an interactive instruction manual for a thermostat by using only the holobuilder platform. Second, we have created a project for monitoring a machine used in an assembly line for gears based on OPC UA.

5.1 Interactive Instruction Material

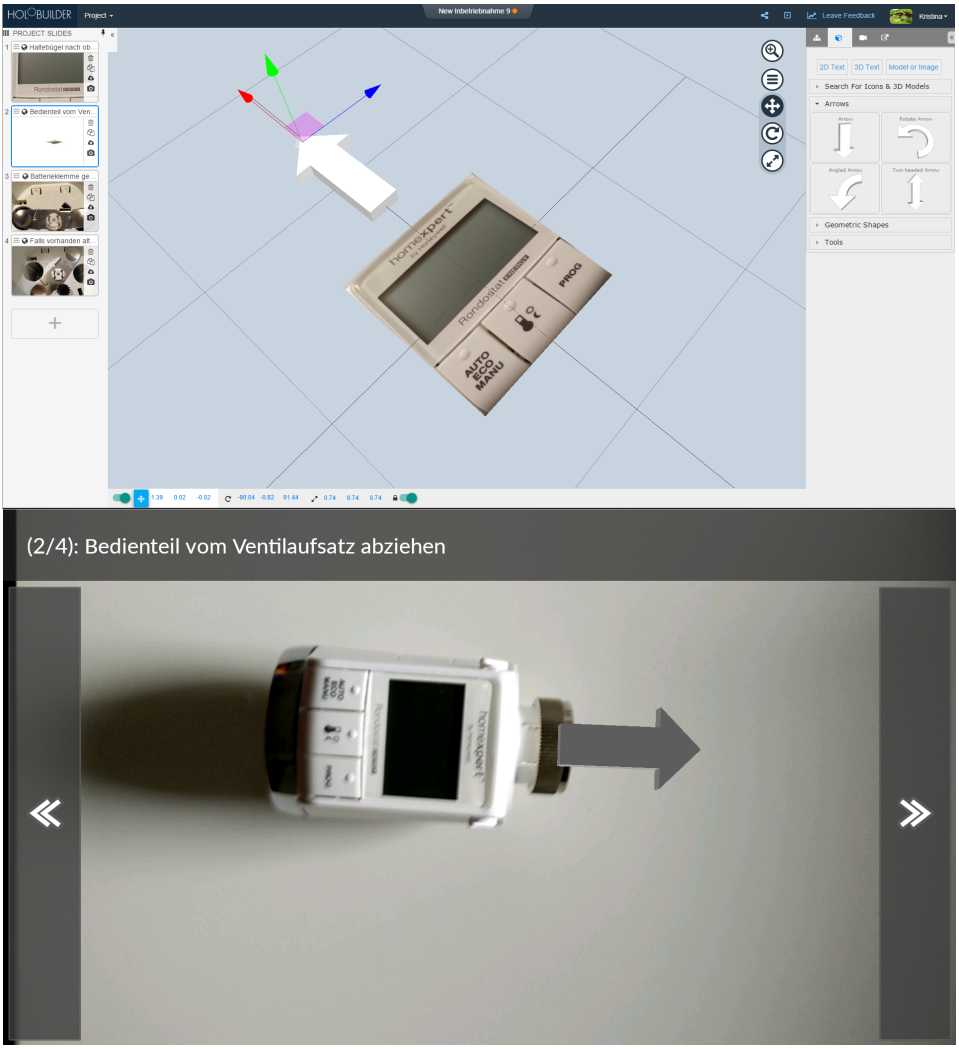


Fig. 4: Interactive instruction manual as created in holobuilder (top) and the corresponding AR scene as viewed through the camera of a smartphone (bottom).

In the context of her Master's thesis, Kristina Tenhaft has created an interactive instruction manual for the thermostat *HR-25* by *Honeywell*[19]. The overall goal is to completely replace the paper manual provided by Honeywell with the interactive instruction manual implemented with holobuilder and to investigate the difference between the two, as perceived by real users (i.e., the core research question is "Can the paper manual be replaced by the AR manual?"). The scenarios covered by the interactive manual include setting date and time, changing batteries, adjusting the temperature, defining heating intervals and opening/closing the vent of the thermostat, among others.

When using the manual with their smartphone or tablet PC, the user first has to select the scenario they want to accomplish. Subsequently, they have to ensure that the thermostat is in the field of view of the camera of their mobile device. For the interactive instruction manual, a 2D marker has been used in terms of a photo of the thermostat surface (cf. Sec. 3 and Fig. 4). Based on this marker, holobuilder recognizes the real thermostat and presents the user with the necessary steps to accomplish their task, which is illustrated by, e.g., virtual arrows (Fig. 4) and other 3D objects. Besides holobuilder's rooms, the manual also makes use of text and checklist slides. In particular, a text slide is used to include a video explanation of how to install the thermostat on a radiator[20]. This is possible since text slides in holobuilder also allow the embedding of arbitrary HTML content. Moreover, checklist slides are used to, e.g., ensure all necessary components have been unpacked and are ready to use.

According to Ms. Tenhaft the implementation of an "average scenario"—i.e., a scenario like changing batteries, which contains several steps and markers—requires about 20 minutes to create for a user who has been previously familiarized with holobuilder. Moreover, she appreciated the possibility of cloning slides. She described this functionality as "highly convenient" since interactive instruction manuals contain a considerable amount of redundant steps that are omitted in paper manuals[21]. Furthermore, Ms. Tenhaft noted that most of the time effort while using holobuilder has to be accounted to creating custom animations (which are "compose[d] from the basic 'rotate', 'scale' and 'position' animations" [Sp15]) due to occasional technical flaws.

5.2 OPC UA

To demonstrate holobuilder's capability of leveraging machine-to-machine communication via OPC UA, we as well implemented a corresponding case study.

The hardware set-up was a machine that used multiple sensors to measure, move, grab and release gears on an assembly line. The contained sensors detected the type of the material and counted the number of gears that passed the assembly line. Moreover, the machine

[19] `http://www.homexpertbyhoneywell.com/en-DE/Products/rondostat/Pages/HR-25.aspx` (accessed 2015/04/27).

[20] This step has been included in terms of a video since it would otherwise require a more complex AR set-up with a variety of different markers or a true virtual reality (VR) scenario. Yet, the latter two have no significant added value compared to the video.

[21] For instance, changing batteries could be a prerequisite for a different scenario.

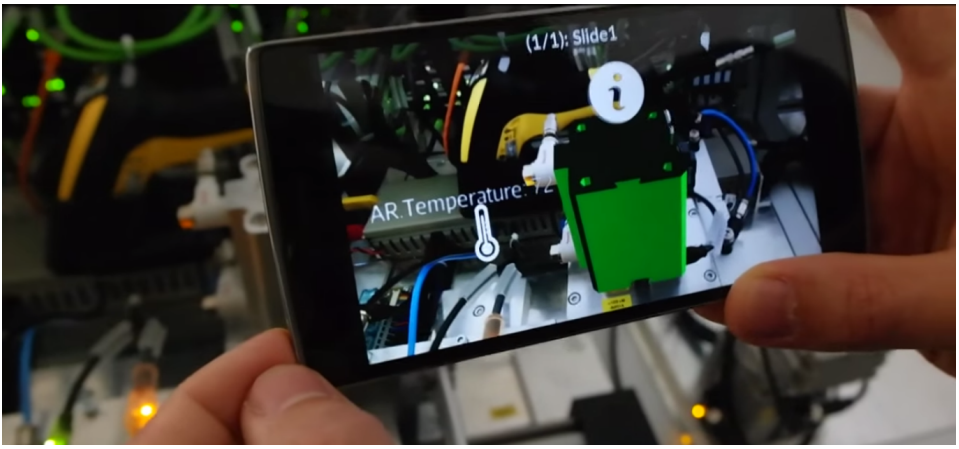


Fig. 5: Case study showcasing real-time information communicated via OPC UA in an AR scenario.

detected and punched out the inner cylinder of the gear if one was present and collected this data as well.

The connection established via an OPC UA server provided all data measured by the sensors to the mobile device displaying the OPC UA–enabled holobuilder project. This included the current temperature of the punch press in case the press was currently working and whether it operated correctly (Fig. 5). The assembly line sensors measured the type of the gear, the material it was made of, whether it had an inlet and the current count of all gears running through the assembly line. Modifying the machine state through the holobuilder project was possible as well. In this case study, we added a virtual button to reset the gear counter of the machine. Hence, we were able to demonstrate the changing of machine properties through our mobile device via OPC UA.

The AR scenario displayed by the holobuilder player (Fig. 5) showed all relevant information on top of the corresponding physical counterpart (i.e., the machine in the real world as viewed through the camera). In this way the information was visible where it was most relevant and the relation between the displayed information and the functionality of the machine became most clear. Additional information could be shown in a separate UI accessible through info buttons “floating” above the machine (Fig. 5). Finally, a virtual version of the punch press was overlaid on top of the real pump and colored in red or green to indicate whether it was currently in use or not (Fig. 5).

6 Discussion & Conclusions

Currently, we are witnesses of the fourth industrial revolution, which has been coined *Industry 4.0* by the German federal government [HPO15, Hi15]. This revolution is driven by various technologies that enable communication between humans, machines and products to form smart factories [HPO15]. One important concept of Industry 4.0 are cyber-physical

systems [HPO15], an example of which are *augmented reality* applications that provide a fusion of the real and the virtual world (which is a property of CPS according to [Ka14]). Yet, to date there are only individual, industry-specific solutions for such AR applications. At the same time it is rather difficult for end-users without specific training to easily create corresponding AR content. Hence, we have introduced *holobuilder*—a novel web platform intended for end-user design of AR content. holobuilder allows for the creation of interactive instruction manuals in Industry 4.0 settings. Additionally, its projects can be enhanced by adding OPC UA capabilities for direct communication between devices displaying AR content and the monitored machines. We have shown that our platform enables Industry 4.0 according to the six design principles specified by [HPO15]. Moreover, its effectiveness and feasibility has been demonstrated based on two real-world case studies involving a thermostat and a machine handling gears in an assembly line.

Yet, our approach still has several shortcomings. First, not all included features are fully mature yet. For instance, as has been stated in one of the case studies, creating custom animations for 3D objects takes a considerable amount of time, which can be attributed to non-optimal intuitiveness and occasional technical flaws. Second, as holobuilder has only recently left the alpha testing state, some crucial functionalities are not yet existent. In particular, our platform still lacks an undo function as well as functionality to copy/-paste 3D objects across rooms. Third, the sets of predefined 3D objects and animations are still rather limited, which we intend to tackle by growing our database and letting users contribute to the publicly available models/animations. In this way it would be possible to provide a "clip art gallery for 3D content" alongside holobuilder. Finally, the holobuilder UI can not be considered optimal yet, as end-user design of 3D content is not a trivial task and has to be communicated to average users in a simple and understandable fashion. Therefore, we are performing ongoing qualitative user studies. Yet, a quantitative evaluation—also in relation to similar approaches—still has to be carried out. Tackling the above shortcomings is our main focus for future work. Particularly, we are confident to be able to report on quantitative findings concerning the effectiveness and efficiency of holobuilder in the near future.

References

[Bi14] Bievor, Marx: Generic OPC UA Modules And Use-Cases For Head-Mounted-Displays. 2014. Seminar Paper, RWTH Aachen University.

[He15a] Hein, Treena: Virtual Fix – Augmented reality tech by B.C.'s NGRAIN takes CMMS to the next level at Lockheed Martin. Design Engineering March/April 2015, 2015.

[He15b] Heinen, Simon: HMI 4.0 – Tablet war gestern [HMI 4.0 – Tablet PCs are a Thing of the Past]. IEE Elektrische Automatisierung + Antriebstechnik 05/2015, 2015.

[Hi15] Hightech-Strategie der Bundesregierung [High-Tech Strategy of the German federal government], `http://www.hightech-strategie.de/`, accessed 2015/04/22.

[HPO15] Hermann, Mario; Pentek, Tobias; Otto, Boris: Design Principles for Industrie 4.0 Scenarios: A Literature Review. 2015. Working Paper No. 01/2015, Audi Stiftungslehrstuhl Supply Net Order Management, TU Dortmund.

[Ka14] Kagermann, Henning: Chancen von Industrie 4.0 nutzen [Taking the Chances of Industry 4.0]. In (Bauernhansl, Thomas; ten Hompel, Michael; Vogel-Heuser, Birgit, eds): Industrie 4.0 in Produktion, Automatisierung und Logistik [Industry 4.0 in Production, Automation and Logistics], pp. 603–614. Springer, 2014.

[OP15] What is OPC?, https://opcfoundation.org/about/what-is-opc/, accessed 2015/04/23.

[Pl15] Plattform Industrie 4.0: Normen und Standards [Platform Industry 4.0: Rules and Standards], http://www.plattform-i40.de/themen/normen-standards, accessed 2015/04/22.

[PS12] Petersen, Nils; Stricker, Didier: Learning task structure from video examples for workflow tracking and authoring. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR '12). IEEE, pp. 237–246, 2012.

[RE15] REFLEKT ONE Augmented Reality Software, https://one.re-flekt.com/augmented-reality-software/, accessed 2015/04/28.

[Sp15] The Definitive Guide to HoloBuilder, https://medium.com/@maxspeicher/the-definite-guide-to-holobuilder-3b62a54d303e, accessed 2015/04/23.

[tHO14] ten Hompel, Michael; Otto, Boris: Technik für die wandlungsfähige Logistik - Industrie 4.0. In: 23. Deutscher Materialfluss-Kongress. 2014.

13. Workshop Automotive Software Engineering

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

13. Workshop Automotive Software Engineering

Heiko Dörr[1] Steffen Helke[2]

Der Trend zum Einsatz immer komplexerer Softwarefunktionen im Fahrzeug ist ungebrochen. Ebenso wächst der Bedarf an immer ausgereifteren Softwarewerkzeugen für die unterstützenden Entwicklungsprozesse. Zukunftsweisende Konzepte für Mobilität und Logistik werden durch Software-Ökosysteme realisiert, deren Funktionalität in zunehmenden Maße auf einer intelligenten Kommunikation mit der Umgebung basiert.

Durch das stärkere Einbeziehen der Umwelt entstehen einerseits neue Anwendungsmöglichkeiten, andererseits aber auch Gefahren, weil die Systeme offener und damit angreifbarer werden. Folglich sind noch mehr Anstrengungen notwendig, um auch weiterhin alle essentiellen Qualitätsmerkmale wie Zuverlässigkeit, Sicherheit (Security und Safety), Bedienbarkeit aber auch den Datenschutz (Privacy) garantieren zu können. Hinzu kommt, dass die Systementwickler mit kontinuierlich wachsenden ökologischen Anforderungen konfrontiert werden. So nutzen Funktionen wie ,,eco-roll" oder ,,E-horizon" die Kenntnis von Topologie und Umgebungsbedingungen zur verstärkten Reduktion des Energiebedarfs und damit des CO2-Ausstoßes.

Der Austausch und die Diskussion darüber, wie neue Funktionen eines in die Umwelt eingebetteten Fahrzeugs bei Erhalt der bestehenden Qualitätsanforderungen umgesetzt werden können, stellen Inhalt und Ziele des diesjährigen Automotive Software Engineering Workshops dar. Diese thematische Ausrichtung bietet viele Anknüpfungspunkte zur INFORMATIK 2015, die mit dem Tagungsmotto ,,Informatik, Energie und Umwelt" ein ideales Umfeld bietet. Der Workshop richtet sich gleichermaßen an Forscher, Entwickler und Anwender aus der Automobilindustrie sowie an Wissenschaftler aus Forschungsinstituten und Hochschulen, die im Gebiet Automotive Software Engineering arbeiten. Im Fokus stehen traditionell weniger theoretische, als vielmehr praxisnahe Arbeiten.

Es wurden 15 Beiträge eingereicht und für jeden Beitrag vier Gutachter festgelegt. Nach Auswertung der Gutachten konnten 10 Papiere zur Veröffentlichung akzeptiert werden. Herzlichen Dank an alle Gutachter, die sich mit viel Engagement in den Begutachtungsprozess eingebracht haben.

Wie schon in den vergangenen Jahren wird das Workshop-Programm mit einer industriellen und einer akademischen Keynote abgerundet. Wir bedanken uns bei Dr. Markus Bechter (BMW AG) und bei Prof. Dr. Jean-Pierre Seifert (Technische Universität Berlin), die beide spontan zugesagt haben und zum Gelingen dieses Workshops entscheidend beitragen.

[1] Model Engineering Solutions GmbH, doerr@model-engineers.com
[2] Brandenburgische Technische Universitt Cottbus-Senftenberg, steffen.helke@b-tu.de

Programmkomitee

Dr. Christian Allmann	Audi AG
Prof. Dr. Marcel Baunach	Technische Universität Graz
Prof. Dr. Manfred Broy	Technische Universität München
Dr. Stefan Bunzel	Continental AG
Dr. Mirko Conrad	samoconsult GmbH
Prof. Dr. Werner Damm	Universität Oldenburg
Dr. Peter Dencker	ETAS GmbH
Dr. Heiko Dörr	Model Engineering Solution GmbH
Bernd Frielingsdorf	Ford Werke GmbH
Prof. Dr. Bernhard Hohlfeld	Technische Universität Dresden
Prof. Dr. Stefan Jähnichen	Technische Universität Berlin / FZI
Ralf Kalmar	Fraunhofer IESE
Prof. Dr. Stefan Kowalewski	RWTH Aachen
Prof. Dr. Thomas Kropf	Robert Bosch GmbH
Prof. Dr. Reiner Kriesten	Hochschule Karlsruhe
Peter Manhart	Daimler AG
Prof. Dr. Klaus Pohl	Universität Duisburg-Essen
Prof. Dr. Wolfgang Pree	Universität Salzburg
Dr. Alexandre Saad	BMW Group
Prof. Dr. Eric Sax	Karlsruhe Institute of Technology (KIT)
Prof. Dr. Ina Schaefer	Technische Universitt Braunschweig
Prof. Dr. Jörn Schneider	Hochschule Trier
Dr. Sebastian Siegl	Audi Electronics Venture GmbH
Dr. Reinhard Stolle	BMW Car IT GmbH
Claus Stellwag	Elektrobit Automotive GmbH
Prof. Dr. Michael Uelschen	Fachhochschule Osnabrck
Dr. Marcel Wille	Volkswagen AG
Dr. Fabian Wolf	Volkswagen AG
Dr. Dirk Ziegenbein	Robert Bosch GmbH

Organisation

Prof. Dr. Steffen Helke	Brandenburgische Technische Universität Cottbus-Senftenberg
Dr. Heiko Dörr	Model Engineering Solutions GmbH

Die Organisation erfolgte in enger Abstimmung mit der GI-Fachgruppe Automotive Software Engineering[1], die den ASE-Workshop seit vielen Jahren veranstaltet.

[1] http://fg-ase.gi.de/

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Softwareplattform für zukünftige automotive Anwendungen

Markus Bechter[1]

Abstract: Vor mehr als zehn Jahren ging AUTOSAR mit dem ambitionierten Versprechen an den Start, eine offene und standardisierte Architektur für Fahrzeugsoftware zu schaffen und auf breiter Basis zu etablieren. Heute lässt sich sagen, dieses Ziel wurde erreicht. AUTOSAR erfüllt die Anforderungen der steigenden Komplexität automobiler E/E-Architekturen und ist zum Standard für Automotive-Software geworden. Die Plattform ist für OEMs und Zulieferer die Basis der Zusammenarbeit mit ihren Partnern und die Grundlage vieler erfolgreicher Projekte.

Zukünftige Anwendungen wie das hochautomatisierte Fahren erfordern jedoch einen hohen Grad an Ausfallsicherheit kombiniert mit sehr hohen Anforderungen an die Rechenleistung und den Speicherplatz. Im Zusammenhang mit der Vernetzung des Fahrzeugs ermöglichen es neuartige Car-to-X-Technologien, wie im Bereich der Consumer Elektronik schon lange üblich, Funktionen im Fahrzeug als eigenständiges Softwareprodukt anzubieten und bei Bedarf nachträglich zu installieren. Diese und weitere sich abzeichnende Anwendungen, genannt sei an dieser Stelle die Vernetzung mit der Infrastruktur, bedürfen der Integration in neue oder bestehende IT-Cloud-Architekturen beziehungsweise in Off-Board-Systeme, wo die Daten an einen Server des Anbieters geschickt werden. Diese lassen sich mit einer neuen Softwareplattform deutlich effizienter umsetzen, weshalb die AUTOSAR Partnerschaft mit der Spezifikation einer zweiten Softwareplattform begonnen hat. Dies wird die neue AUTOSAR Adaptive Platform sein.

Die neue Plattform orientiert sich in der Softwarearchitektur sowie in Bezug auf die Hardwareanforderungen deutlich stärker an Lösungen, die im Consumer- und Telekommunikationsumfeld bereits weit verbreitet sind. Bei der Spezifikation der AUTOSAR Adaptive Platform soll daher auch verstärkt auf existierende Technologien und Standards wie beispielsweise den Unix API-Standard POSIX (Portable Operating System Interface) zurückgegriffen werden. Dabei ist natürlich sicherzustellen, dass diese Technologien den besonderen Anforderungen der Automobilindustrie gerecht werden und entsprechend erweitert werden.

Der Vortrag stellt die wesentlichen Aspekte und Herausforderungen bei der Definition einer neuen Standardplattform für automotive Software dar und zeigt den Weg auf, wie diese Plattform reibungsfrei in die existierenden Systeme integriert werden kann.

[1] BMW AG, DE, markus.bechter@bmw.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Safety Functions on Commodity Hardware with Diversified Encoding

Martin Süßkraut[1], André Schmitt[2] and Jörg Kaienburg[3]

Abstract: Currently, hardware designed and certified for safety-critical systems is one important building block for any safety-critical application. Such hardware provides the detection of execution errors. However, many modern safety-critical applications, like autonomous driving, require features and performance levels that are not available from safety-certified hardware. One solution to this problem is to use hardware that is not certified for safety-critical systems, for example consumer-graded hardware, but that fulfills the feature and performance requirements. Additionally, a software solution provides the detection of execution errors.

This paper introduces such a software solution called "Diversified Encoding with Coded Processing". Due to its hardware-independence, this solution provides the flexibility to build safety-critical systems from non-safety-critical hardware components. This solution can be automated with a code transformation tool to further increase the flexibility.

Keywords: safety, ISO 26262, software coded processing, software diversification

1 Introduction

Traditionally, safety-critical hardware is designed for detecting execution errors. Execution errors are errors that falsify the execution of a correct software. Examples are transient bit-flips in memory, CPU or on the BUS and permanent errors like stuck-at 0 or 1 on gates in the ALU of the CPU. Current non-safety-critical hardware or consumer-graded hardware, respectively, are typically unable to detect execution errors but provide significantly higher performance. Hence, safety-critical systems that require high performance or special features that are only available in consumer-graded hardware, e.g. GPUs, need a new solution for detecting execution errors.

This paper gives an introduction into the software solution called "Diversified Encoding with Software Coded Processing": A solution to continuously detect execution errors at runtime and allowing to make any system intrinsically safe. Because, diversified encoding is hardware-independent it can be used with consumer-graded hardware. After the introduction in Section 1, Section 2 presents an example which will be used during the paper. Sections 3 and 4 introduce software coded processing and diversified encoding, respectively.

ISO 26262 references coded processing based on Forin [Fo89] in Volume 5 under D 2.3.6 as a technology to detect hardware errors [IS11]. Forin is also the basis for software coded

[1] SIListra Systems GmbH, Königsbrückerstr. 124, 01099, Dresden, martin.suesskraut@silistra-systems.com
[2] SIListra Systems GmbH, Königsbrückerstr. 124, 01099, Dresden, andre.schmitt@silistra-systems.com
[3] SIListra Systems GmbH, Königsbrückerstr. 124, 01099, Dresden, joerg.kaienburg@silistra-systems.com

processing introduced in Section 3 [SSF09]. Software coded processing in combination with diversified encoding is a flexible solution to detect execution error with the potential to fulfill ASIL-D. ASIL-D is the highest safety level of the ISO 26262. The utilization of diversified encoding via a fully automated code transformation tool allows the efficient deployment of diversified encoding across different industries, e.g. automotive, medical, automation, rail, and avionics.

Section 5 discusses a code transformation tool for generating the diversified encoding solution. To achieve this, the tool takes the source code of a safety function as its input. The output of the tools is again source code. This generated source code provides the same functionality as the given safety function plus the capability to detect hardware errors.

The novelty of the presented tool is the broad support for the C programming language. The diversified encoding solution automatically safe-guards against transformation failures of the transformation tool. Previous approaches using coded processing do not have these properties.

The paper concludes in Section 6 with a discussion of the advantages of the diversified encoding approach: hardware-independence, flexibility of software and tool support for the C programming language. Hence, development and innovation cycles could be expedited. As one of the intended side effects, diversified encoding decouples hardware-error detection and functionality. The presented technique of diversified encoding with coded processing allows system architectures equipped with safety functions while running on non-certified hardware.

2 Example

The following exemplary safety function will help to explain and demonstrate the diversified encoding and its properties and advantages. This example focuses on the overall approach of diversified encoding and code generation.

The example safety function is a counter, that can be used as a reference to check a sequence of received messages. It has the following specification:

- The counter is one 32-bit counter unsigned integer value.
- The initial counter value is 1.
- The counter has an interface to query the current counter value and to increment the counter value by a given unsigned integer value.
- Execution errors while executing the counter must be detected.

For a real world application, one needs to extend this specification to include aspects like integer overflow and concurrency.

The software developer designs a C interface from this specification:

```
uint32_t incSafeCounter(uint32_t incValue).
```

The implementation itself does not contain any specific measures to detect execution errors. Sections 3 and 4 demonstrate how to automatically generate the source code to detect execution errors. Especially Section 4.1 shows how to change the interface to enable the detection and handling of execution errors. The following source code shows a possible implementation of this safety function (with C99 fixed-width integer types[4]):

```
1  /* include standard C library for fixed-sized integer types */
2  #include <stdint.h>
3
4  /* the safety critical counter variable */
5  static uint32_t safeCounter = 1;
6
7  /* interface of the counter */
8  uint32_t incSafeCounter(uint32_t incValue) {
9    safeCounter += incValue;
10   return safeCounter;
11 }
```

3 Software Coded Processing

Diversified encoding builds on the coded processing technology [Fo89]. Software coded processing (SCP) adds information redundancy to a software program to enable it to detect execution errors. To integrate SCP into a software program one has to either rewrite the program manually or one can use an automated transformation tool. SCP is applied to the complete data flow of a program. The data flow of a program is the sum of all computations and data processing in the program. In other words, all constants, variables and operations have to be re-programmed while making use of SCP.

The literature describes different encodings. The most prominent encoding is the AN encoding: Every value in the program is a multiple of a constant A [SSF09]. AN is the name of the encoding and not an abbreviation. Values that are not multiples of A are considered as invalid. With SCP, all operations in a program must work with these encoded values. An execution error produces invalid values.

A helpful example is the summation $2+3$. Encoding with $A=7$, the calculation turns into $14+21$. Without an execution error, the result is 35 and a valid value because it is a multiple of 7. The criterion which decides whether the result is valid or invalid is the property that valid results are a multiple of A. If not, results are considered as invalid.

Two principle kinds of execution errors can influence the result:

- One of the input values was already an invalid value. This can either happen because of a bit-flip in the memory, that holds the value, or when the value was already the result of an erroneous computation. For example, if the 14 changes into a 13, the result is 34 which is not a multiple of 7.

[4] The type `uint32_t` is a 32-bit unsigned integer. The type `uint64_t` is a 64-bit unsigned integer.

- The operation itself can be incorrect. For example, the summation could add an additional 2 to the result. Then, the result is 37, which is not a multiple of 7 and, again, not a valid value.

An encoded program can check any encoded variable at any point in time. However, for performance reasons it is better to only check the output values (see Section 4.2). In-between checks are not necessary because the error propagates through the data flow of the program.

The source code below shows the example from Section 2 with a simplified AN encoding for $A = 7$:

```
1  /* include standard C library for fixed-sized integer types */
2  #include <stdint.h>
3
4  /* 1 is AN encoded 7 */
5  static uint64_t safeCounter = 7;
6
7  uint64_t incSafeCounter_encoded(uint64_t incValue)
8  {
9    safeCounter = add_encoded(safeCounter, incValue);
10   return safeCounter;
11 }
```

The main differences between the native safety function and the encoded safety function are:

Data types The data types change from `uint32_t` to `uint64_t`. SCP needs additional bits for the information redundancy. The encoded version of the largest 32-bit unsigned integer value does not fit into 32-bit[5].

Constants All constants must be AN encoded. In the example, the initialization value of `safeCounter` is now 7 – the AN encoded 1.

Operations Encoded operations replace all native operations. Encoded operations operate on encoded values like native operations operate on native values. Hence, encoded operations must also implement over- and underflows, where they are defined by the C-standard. For instance, the `add_encoded` returns the encoded sum of its two encoded input values.

The safety of the AN encoding depends on the constant A. The following properties of A influence the safety of AN encoding:

Size The larger A the more safe is the encoding. If n is the number of bits required to represent all native values (e.g. 32-bit) and k is the number of bits to represent A (e.g. 3 bits for $A = 7$) $n+k$ bits are needed to represent all encoded values. The number of valid encoded values is the same as the number of all native values: 2^n. Only these 2^n out of all words representable with $n+k$ bits are valid words. All other words are invalid. The probability for a completely random error to change a valid

[5] The largest 32-bit unsigned integer value is $4,294,967,295$. Encoded with $A = 7$ this value is $30,064,771,065$.

word into another valid word is roughly $\frac{2^n}{2^{n+k}} = \frac{1}{2^k}$. The probability $\frac{1}{2^k}$ depends only on the size of A.

Hamming Distance Because computers represent values as binary words, the hamming distance of the AN encoding is also important. The hamming distance depends on the value of A. For example, A should never be a power of 2, because the resulting AN encoding has the hamming distance of 0. The reason is that multiplying with a power of two is the same as a shift.

4 Diversified Encoding

Diversified encoding is based on two distinct executions of the same safety function. These two executions are:

Native Execution: The native execution is the result of the original source code of the safety function. This source code operates on the native input values. This execution only changes native state. The result of this execution is the native output.

Encoded Execution: The encoded execution is based on the encoded variant of the safety function. This execution operates on encoded input values and on the encoded state. It produces an encoded output.

Both executions are completely distinct computations but operate on the same values. The encoded input values are the encoded variants of the native input values. The source code of the original, native code is used to generate the encoded source code thereof. This can be done either manually or, recommended due to the high degree of reproducibility, via a code transformation tool.

A diversity framework manages both executions. The transformation tool generates the source code of the diversity framework from the source code of the original safety functions.

The component that uses the safety function can detect and handle execution errors with the help of checksums. We call such a component a caller. The diversity framework generates two checksums: one over the native output values and another over the encoded output values. A caller operates solely on the native input and output values. After executing the safety function, a caller must compare the checksum over the native and encoded output values to verify whether the native output values are from an error-free execution. Section 4.2 shows an example for such a check. If these checksums differ, an execution error has happened. A standard checksum algorithm can be used for computing the checksum, e.g. CRC32.

Fig. 1 shows the dataflow of a safety function with the diversified encoding solution. The data flow starts at ① with the native input state. The native input state is protected by a checksum. A caller must calculate this checksum, as soon as the caller has assembled the input state. When a caller executes the safety function, it passes the control to the diversity framework. In step ①, the diversity framework encodes and checks the native input values. The result of step ① is the encoded input values for step ②. The diversity framework uses

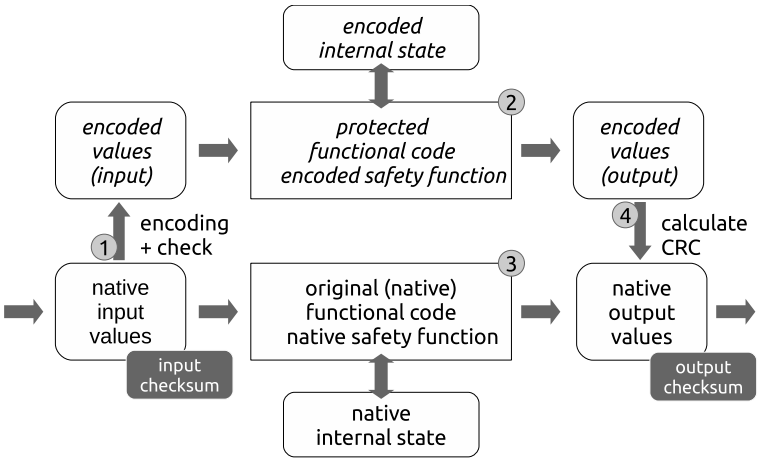


Fig. 1: Data-flow model of diversified encoding.

the input checksum over the native input values to check the correctness of the encoded input values. Next, in step ②, the diversity framework executes the encoded safety function. This function reads the encoded input values and the encoded internal state. It performs its calculations, updates the encoded internal state and produces the encoded output values. In step ③, the diversity framework executes the native safety function F_N. Function F_N reads the native input values. F_N operates on the native internal state and produces the native output values. In the last step ④ the diversity framework calculates the checksum of the native output values over the encoded output values. Then the diversity framework passes the control back to the caller. It's the caller's responsibility to check the checksum of the native output values.

The safety function works with the following parts:

- native input values
- input checksum
- native safety function
- diversity framework (including the encoded safety function) generated by the code transformation tool
- native output values
- output checksum

The code generation tool generates the source code of the steps ①, ② and ④. Steps ① and ④ are part of the diversity framework. Step ② is the encoded safety function. The source code of step ③ is the original source code of the native safety function.

The solution detects transient execution errors because of the two executions. Permanent errors are detected, because of the diversity between the two executions. The different data representations between the two executions enable to detect freedom of interference errors

with respect to memory as execution errors. Timing errors can be detected with a watchdog (see Section 4.2). For a more detailed the safety argument see [SSK15].

4.1 Diversified Counter

This section explains the usage of diversified encoding with the example from Section 2. The code transformation tool generates an entry point function `incSafeCounterSafe` for the function `incSafeCounter`. This entry point is part of the diversity framework generated by the code transformation tool and it encapsulates the four tasks from Fig. 1:

1. It generates the encoded input values from the native input values.
2. It performs the native execution.
3. It performs the encoded execution.
4. It calculates the checksum over the encoded output.

The new entry point has the following interface:
`uint32_t incSafeCounterSafe(uint32_t* checksum, uint32_t incValue).`
The new entry point has a new name, because the native entry point `incSafeCounter` is still used in step ③. The return value and the parameter `incValue` have the same semantics as in `incSafeCounter`. A caller must provide a pointer to the `checksum` variable[6]. The new entry point `incSafeCounterSafe` stores the checksum over the encoded output into this `checksum` variable.

To check for execution errors, e.g. when using the return value of `incSafeCounterSafe`, the caller's code must compare the value of the variable checksum with the checksum of the native output values. To calculate the checksum over the native output values, the transformation tool generates the function:
`uint32_t SIListra_diversity_output_checksum(uint32_t returnValue).`
The parameter `returnValue` must be the return value that `incSafeCounterSafe` returns. We exemplify the usage of the output checksum below with an integration of a watchdog.

4.2 Integration with a Watchdog

This section shows how to integrate diversified encoding with a challenge-response-watchdog (CRWD). A CRWD monitors a safety system. It regularly sends challenges to the safety system and expects a correct response within a defined time frame. In case the response is wrong or does not arrive in a timely manner, the CRWD initiates the appropriate safety reaction for the system, e.g. a stop, reset, or fail-over. Such CRWDs are state-of-the-art in the industry.

Challenges and responses are usually integer values. For each possible challenge, the expected response has to be pre-calculated. These expected responses are stored within the CRWD as part of the configuration of the system.

[6] The `checksum` variable can also be global variable, if pointer usage is discouraged.

Our goal is to show how to enable the CRWD to monitor, that (1) the safety function is regularly executed and (2) that the executions are free of execution errors. The source code below shows how diversified encoding can generate a response to a given CRWD challenge. In the following discussion, any challenge and response shall be represented as unsigned 32-bit integer variables. The CRWD writes the variable challenge whenever it sends a new challenge. In addition, the CRWD clears the variable response, e.g. by setting it to an invalid response value. When the CRWD expects a response, it reads the variable response and compares its value to the expected response. In practice, the value of response can be merged with responses from other safety functions to calculate a unified response for the whole system.

```
1  /* constant for calculating the response for the counter */
2  #define SAFE_COUNTER_SIG   (12u)
3
4  /* stores the current watchdog challenge */
5  extern uint32_t challenge;
6  /* stores the response for the watchdog */
7  extern uint32_t response;
8
9  /* initialize checksum variable with unlikely checksum value */
10 uint32_t checksum = 0;
11
12 /* increment the counter by 1 */
13 uint32_t currentCounterValue = incSafeCounterSafe(&checksum, 1u);
14
15 /* calculate response from challenge and output checksum */
16 response = challenge + checksum + SAFE_COUNTER_SIG -
17     SIListra_diversity_output_checksum(currentCounterValue);
18 /* ... */
```

When no execution error influenced the output value `currentCounterValue`, the values of `checksum` and `SIListra_diversity_output_checksum(currentCounterValue)` are the same. The expected response for a given challenge is `challenge + SAFE_COUNTER_SIG`.

If the values of `checksum` and `SIListra_diversity_output_checksum(currentCounterValue)` differ, then an execution error happend. In this case the value of `response` will deviate from the expected response value. Even when the safety function is not executed at least once between a new challenge and its corresponding time for a response, the variable `response` contains an invalid response value. Hence, the CRWD detects whether the safety functions is regularly executed and whether an execution error influences the execution of the safety function.

The presented solution always overwrites the variable `response`. Hence, a fault-free execution of the safety function might mask a previous error in case the CRWD did not check `response` between these two executions. To avoid the masking of a previous error one can extend the solution to propagate an execution error through consecutive executions into the response that the CRWD reads.

5 Code Transformation

The code transformation tool generates automatically the source code of the encoded program and the source code of the diversity framework. The diversified encoding solution safe guards against tool errors, which positively influences tool classification. In case the code transformation tool generates source code for the encoded safety function that does not behave as the original source code, the output values of native and encoded safety function do not match. Hence, diversified encoding per se detects any wrong output that results from an in-correct generated encoded safety function as execution error.

The code transformation tool works like a C compiler. It performs the following steps in the shown order:

Preprocessing The tool preprocesses the complete source code of the safety function including all included header files. Therefore, it must be configured with the same preprocessing defines as the C compiler.

Parsing The tool generates an abstract syntax tree from the preprocessed source code including a semantic analysis. For instance the abstract syntax tree needs a complete type analysis. At this step, the tool can detect unsupported C languages features. If the tool does not support a language feature, it produces an error message and aborts the transformation process.

Encoding Sufficient semantic information is available after the parsing step to encode the safety function. The tool replaces constants, variables, and operations with their encoded versions.

Code Generation The result of the encoding step is an intermediate result. The last step is the final C code generation.

The current status of the tool allows to encode most C99 features. The following features are known to be supported by a state-of-the-art code transformation tool:

- all integer arithmetic including logical operations and comparisons
- all control flow constructs of C99 from if-else, including loops, function calls to break and continue (except setjmp and longjmp)
- arrays, structs and any other pointer arithmetic

The encoding of floating point arithmetic is still a research topic.

For each native C module that the tool gets as input, it produces an output C module containing the encoded version of the native C module. The generated C code must then be further processed in the tool chain, typically by the C compiler.

Optimizations in the C compiler cannot remove the encoding. The fact that all values are a multiple of an encoding parameter A is not visible to the C compiler because the C compiler processes each C module individually. Even with link time optimization, which provides the optimizer the view on all C modules at once, the optimizations cannot remove the encoding. The optimizations would have to prove that the whole encoded safety function is equivalent to the native safety function, including all checksum calculations in the diversity framework. No state-of-the-art compiler provides such optimizations. In addi-

tion, fault injection could be used to demonstrate that the encoding has not been removed during compilation.

Besides encoding itself, the code transformation tool can also generate the diversity framework. The diversity framework consists of the source code of the functions that calculate the native input checksum and the native output checksum and the source code of the new entry points.

6 Discussion and Conclusion

This section concludes the paper with a discussion of the current limits of diversified encoding and its advantages.

6.1 Scope of Detection

Diversified encoding is a probabilistic solution to detect execution errors with a very high likelihood. It is based on coded processing. Section 3 introduced the important properties for the encoding parameter A to achieve a very high detection probability. In terms of the ISO 26262, experiments and analysis have shown that state-of-the-art diversified encoding reaches a high diagnostic coverage sufficient for the highest safety level of the ISO 26262. For instance, in [Gh15] we show that only 0.002% of 300.000.000 injected errors where not detected with the diversified encoding solution.

Diversified encoding covers the detection of execution errors in all calculations of the safety function, i.e. in its data flow and in the state of the safety function (memory where the variables of the safety function are stored). Input and output can be covered with the solutions introduced in Section 4.

Timing errors, e.g. a slow execution speed or a crash of the safety function, cannot be detected with diversified encoding alone. Section 4.2 showed how to connect a diversified encoded safety function with a watchdog to detect and handle such timing problems reliably.

Systematic software errors, i.e. software bugs, are not covered by diversified encoding. The encoding technology and the code transformation tool have no information about a specification of a given safety function.

Direct hardware accesses, e.g. via assembler code, are outside the scope of an automatic code transformation tool. It is possible to integrate assembler code with the help of manual encoding. In such cases, it can make sense to use a mix of automatic code transformation and manual encoding.

6.2 Advantages of Diversified Encoding

The decision to use diversified encoding for detecting execution errors is a design decision that impacts sensitive parts of a system architecture and of the development process. To get the best results, diversified encoding has to be considered at an early stage in the development process.

Diversified encoding works hardware independent. It makes no assumptions on the hardware. The only requirement is that there exists a standard C compiler for the target hardware. The target hardware does not need to be certified or developed for safety-critical systems. It is even possible to use consumer-graded hardware together with diversified encoding to develop a safety-critical system.

Because diversified encoding is a software solution, it is more flexible than a hardware solution. Diversified encoding can be restricted to the safety-critical parts of a given system and puts no restrictions on the non-safety-critical system parts. An automated code transformation tool further increases the flexibility and frees development resources that can be assigned to developing the safety function itself.

The coded processing is the base of the diversified encoding. It is possible to use coded processing without diversified encoding. However, diversified encoding has two advantages over coded processing:

- Diversified encoding with AN encoding detects the same symptoms then one can detect with the more complex encodings ANB and ANBD [Sc10a] without diversified encoding. Because of their complexities, ANB and ANBD require more computing resources than AN with diversified encoding [Sc10b]. In other words, diversified encoding requires a lower amount of computing resources than coded processing on a comparable safety level.
- Diversified encoding detects tool errors. The code transformation tool generates the encoded safety function. However, because the diversified encoding compares the output of the native safety function with the output of the encoded safety function, it detects errors of the code transformation tool that alter the functionality of the encoded safety function. Hence, the tool criticality is below the tool criticality of a compiler.

Especially when using the automated code transformation tool, the diversified encoding reduces the development effort for the final system. In addition, the diversified encoding makes cyclic memory checks and periodic instruction set tests for the safety function obsolete. Inverse storage of variables does not need to be applied to variables stored within the safety function. By avoiding these defensive programming methods, valuable development resources are released by the use of coded processing and diversified encoding. And, because these programming methods also consume system resources at runtime, their avoidance releases hardware resources of the safety-critical system.

In conclusion, diversified encoding with SCP solves the problem of the unavailability of feature-rich and powerful hardware that is certified for use in safety-critical systems. Di-

versified encoding with SCP enables the use of modern hardware that is not certified for safety-critical systems such as many ARM CPUs. The possibility to use such modern hardware in safety-critical systems enables the development of a new generation of safety-critical systems that can be smart, powerful and safe at the same time.

References

[Fo89] Forin, P.: Vital Coded Microprocessor Principles and Application for Various Transit Systems. In: IFA-GCCT. pp. 79–84, Sept 1989.

[Gh15] Ghadhab, Majdi; Kaienburg, Jörg; Süßkraut, Martin; Fetzer, Christof: Is Software Coded Processing an Answer to the Execution Integrity Challenge of Current and Future Automotive Software-Intensive Applications? In: Proceedings of AMAA 2015, 19th International Conference on Advanced Microsystems for Automotive Applications. July 2015.

[IS11] ISO TC22/SC3/WG16: ISO/DIS 26262 - Road vehicles – Functional safety. Technical report, Geneva, Switzerland, July 2011.

[Sc10a] Schiffel, Ute; Schmitt, André; Süßkraut, Martin; Fetzer, Christof: ANB- and ANBDmem-Encoding: Detecting Hardware Errors in Software. In (Schoitsch, Erwin, ed.): Computer Safety, Reliability, and Security. volume 6351 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 169–182, 2010.

[Sc10b] Schiffel, Ute; Schmitt, André; Süßkraut, Martin; Fetzer, Christof: Software-Implemented Hardware Error Detection: Costs and Gains. In: The Third International Conference on Dependability (DEPEND 2010). IEEE Computer Society, Los Alamitos, CA, USA, pp. 51–57, 2010.

[SSF09] Schiffel, Ute; Süßkraut, Martin; Fetzer, Christof: AN-Encoding Compiler: Building Safety-Critical Systems with Commodity Hardware. In: SAFECOMP '09: Proceedings of the 28th International Conference on Computer Safety, Reliability, and Security. Springer-Verlag, Berlin, Heidelberg, pp. 283–296, 2009.

[SSK15] Süßkraut, Martin; Schmitt, André; Kaienburg, Jörg: Safe Program Execution with Diversified Encoding. In: Proceedings of the 13th embedded world conference 2015. February 2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Lightweight Contracts for Safety-Critical Automotive Systems

Bernhard Kaiser[1], Stefan Sonski[2], Suryo Buono[3], Hauke Petersen[4], Justyna Zander[5]

Abstract: Complex automotive systems are composed of subsystems and components in a deep hierarchy, often designed by different development partners or reused from preexisting projects. It is therefore a challenging task to break down requirements into sub-requirements fitting the scope of the subsystems and to simultaneously demonstrate that the integrated system fulfills both functional and safety requirements specified on the top-level. Contract-based development is a popular approach for breaking down requirements onto components by means of assumptions and guarantees. However, most current approaches are based on a formal semantics and therefore limited in their expressive power and their acceptance by practitioners from automotive industries. We propose a semi-formal approach that allows specifying assumptions and guarantees at component interfaces in a language with well-defined syntax, but leaving the verification of fulfillment of the contract by a component to expert decision. However, some of the relevant refinement relations can be formalized and automatically checked. We describe our prototypical Eclipse tool that allows the annotation of components with assumptions and guarantees, and the partial checking of the decomposition. We show the applicability by a case study of an automotive electric drive system.

Keywords: contracts, safety, functional safety, requirements engineering, assumption, guarantee.

1 Introduction

Automotive systems have kept growing in complexity for the past decades, and most automotive embedded systems are safety-related systems. Vehicle functions, for example, in the domains of hybrid/electric powertrain or advanced driver assistance systems (ADAS) are distributed across many electronic control units (ECUs), each of them contributing to many different functions. These ECUs again consist of a complex network of components implemented in different technologies. ECUs or components are usually designed by different departments or external suppliers. Some of them are re-used from past projects or have originally been developed stand-alone for later usage in various vehicle projects by different OEMs ("Safety Elements out of Context" in ISO 26262 language). Often we noticed a lack of understanding which requirements and properties a specific functional unit supposedly fulfills, and under which assumptions it had originally been developed. This makes it difficult for the vehicle OEM to provide a sound safety argument for the integrated system. As the degree of automation in vehicle

[1] Berner & Mattner Systemtechnik GmbH, Gutenbergstr. 15, Berlin, bernhard.kaiser@berner-mattner.com
[2] Alpha EOS, Stuttgart, s.sonski@gmail.com
[3] Berner & Mattner Systemtechnik GmbH, Gutenbergstr. 15, Berlin, suryo.buono@berner-mattner.com
[4] Freie Universität Berlin, Computer Science Institute, Takustr. 9, 14195 Berlin, hauke.petersen@fu-berlin.de
[5] Berner & Mattner Systemtechnik GmbH, Gutenbergstr. 15, Berlin, justyna.zander@berner-mattner.com

constantly grows, functional safety gets more and more intertwined with the correctness, accuracy and performance of the nominal function, so the above considerations are not restricted to safety in a narrow sense but apply to correct development of component-based automotive systems in general. Accordingly, system architects need help with systematically decomposing requirements onto components, evaluating candidate components w.r.t. given requirements and demonstrating that all parts fit each other and the integrated system is proper for the selected purpose. Given the situation in industry today, it may be postulated that even an argument that is not provable by a computer but needs manual review is better than none. The decomposition of properties like value accuracies, time bounds for system reactions or safety integrity levels still requires a lot of scientific work and we therefore doubt that a formal solution in combination with a user friendly interface for industry practitioners will be available in the next few years.

An increasingly popular approach for decomposing requirements onto components and showing their compatibility is the development process by contracts. Contracts support partition of work and isolate properties at interfaces. Contracts have first been proposed for the compositional verification of linear software programs [Me92] but were in later years extended by a large research community towards component-based software, i.e. software components as independently created entities that communicate with other components via interfaces. Today, contracts are often understood as sets of assumptions, on which a component relies to be fulfilled by its environment, and guarantees a component makes towards its respective environment. Both assumptions and guarantees refer to observable behavior at the component interfaces. Often, formal notations are used to express contracts (i.e. some forms of predicate logic, temporal logic or other types of modal and higher-order logic), allowing for refinement correctness check of contracts down the architectural hierarchy.

Although the expressiveness of formal contracts has increased impressively, many approaches are still on the level of academic research and lack direct applicability to automotive industries for several reasons: first, when aiming at a formal underpinning, it cannot be expected that the expressive power of today's contract annotation languages will soon fit the great variety of important properties of component-based automotive systems, encompassing properties such as correct function (in whatever sense), accuracy timing and safety integrity. In particular, the latter is rather a meta-property that cannot be formally proven but needs expert judgment for verification. Second, from their origin in informatics, many of the formalisms are only applicable to software components, not system level components. For example, many of the existing frameworks for contracts are based on discrete–time and/or discrete-state components, whereas the physical world is mainly characterized by continuous-valued systems. Last, automotive systems are designed in cooperation of engineers from different disciplines ranging from software engineering over electronics to mechanical engineering, many of them still reluctant to learn and apply formal languages from computer science, but accepting programming-language style notations or graphical executable notations such as Simulink.

Therefore we came to the idea to try it the opposite way: believing that assumptions and

guarantees are a valuable aid for stating, decomposing and allocating requirements, even if written down in natural language and checked manually, we propose a semi-formal notation for assumptions and guarantees. Leaving aside formal semantics in the beginning, we start by proposing a structured language, checkable for well-formedness and yet acceptable to practitioners. Starting from there, we can go on with adding formal semantics at least for *some* aspects and provide checkers for these aspects. For instance, if we manage to provide a compare function that is able to conclude that an accuracy of 5% is better than 10% or that ASIL B is better than ASIL A, then the validity of an assignment of requirements to a set of given components (often referred to as dominance relation) can be confirmed, even without knowing how properties like ASIL or accuracy could be formally defined. The compliance of an individual component with the guarantees still needs to be verified manually, by manual review, simulation, or testing. The approach works in both cases: the breakdown of requirements onto components to be developed, and the reuse of pre-existing components in new projects, provided that they had been specified by contracts in the past. Still the way is open to add formal semantics for more and more aspects in the future (e.g. the validity of an ASIL decomposition can be checked if the ASIL decomposition rules from ISO 26262 are added, or the achievement of some timing or accuracy bounds can be checked by model checkers or simulation, if the semi-formal notation is enriched by a formal underpinning for *certain* properties). Therefore, we do not see our proposal as a competitor to formal approaches, but rather as a complementary to them, preparing the ground for their acceptance in industry.

In a master thesis at Berner&Mattner Systemtechnik [So13] the approach is exemplified by using a tentative specification language which resembles commonly known object-oriented (OO) programming languages like C++. As a proof of concept, a corresponding prototypical Eclipse-based modeling and checking tool was developed. We applied the approach to a case study of an electric drive that is a simplified but yet realistic abstraction of electric powertrain drive systems we encounter in our day-to-day work with well-known German vehicle OEMs and suppliers. This paper summarizes and explains the experiences we gained with our semi-formal approach.

The rest of the article is structured as follows: Section 2 provides a brief overview on contracts; Section 3 presents our approach and introduces the annotation language for assumptions and guarantees. Section 4 introduces the prototypical tool, Section 5 gives an insight to our case study of an automotive electric traction drive and Section 6 concludes with an outlook on the next steps on the research agenda.

2 Background and State-of-the-Art

2.1 Contract-Based Development in the Domain of Software Engineering

Contracts as a mean to assure correctness of computer programs have initially been

proposed by Bertrand Meyer [Me92]. The idea is to avoid errors in the collaboration of different (sub-) functions of a software program by making the expectations of a subfunction towards its caller and vice versa explicit (e.g. regarding the range of allowed values that some variable may take). Thereby, caller and callee engage in a sort of formal agreement, called contract. The expectations to be fulfilled at the entry point of a subfunction are called preconditions and the expectations at the exit point are called postconditions. Preconditions and postconditions may be statements that evaluate to true or false (e.g. parameter x is greater or equal than zero), which can be stated in a formal language which is intrinsic part of the programming language or available as an add-on (e.g. assert-Macros in C language, Object Constraint Language OCL in UML). The expectation of the architect is that at runtime the expression always evaluates to true when reaching the given point in the program flow.

2.2 Contracts for Component-Based Development of Embedded Systems

Contracts in this original sense are applicable to software programs written in linear programming languages. Software and systems architecture for technical systems, however, is today designed and structured using component-based modelling techniques, such as UML Component Diagrams, SysML Internal Block Diagrams, or the data flow oriented and hierarchical subsystems of Simulink. Software and hardware components may be delivered from different suppliers, purchased as standard components off-the-shelf, or reused from former projects. They are integrated by system suppliers or the OEM (e.g. vehicle manufacturer). OEM releases the embedded system to the market and takes responsibility for the correctness, reliability, and safety of the entire system. It is reasonable to transfer the contract concept to this setting.

Contracts described in related work (for an overview see [Be12]) mostly use the concept of assumptions and guarantees. The contract is fulfilled, if, for the integrated system, every assumption is implied by an appropriate guarantee by some peer component. There have been proposals for the usage of contracts to specify real-time properties of continuous-valued controller structures and the control error of such systems (e.g. [Sc14]). Many times, contracts have been proposed for the functional safety domain ([BHM03][Fe07][Ba10] and others). Contracts have been applied to UML/SysML models as well as Simulink models [Bo11] that play an important role in automotive engineering.

3 Approach to Annotate Components with Interface Contracts

Our approach, first developed in [So13] is built upon *Interface Contracts*, which occur between neighbor components. Their usage for the recursive breakdown of system requirements onto the system's components has been elaborated in [BKW15] (cf. Section 3.3). Interface Contracts are bound to the input and output interfaces (ports) of a component and are engaged between neighbour components that are directly linked by signal flows (producer and consumer of some signal). In our approach, we assume the

system architect to be an omniscient observer. Therefore, it is allowed, for instance, to describe the accuracy of some measurement signal as the maximum expected deviation (e.g. 6 sigma) to its real-world counterpart signal, or the delay of some internal event as being the time passing between some event in the real world (e.g. some vehicle crash occurs) to the internal event (e.g. the command to some airbag ignition unit to inflate).

Our proposed process of requirements refinement and allocation by contracts comprises several steps: first, requirements are specified on system level. These form the guarantees to be fulfilled by the system under development, provided that the assumptions about the system's operational environment hold. Next, the architect proposes a candidate system architecture in terms of components (here understood as functional blocks, independent of their realisation in hardware or software) linked by signal flows via their interfaces, with the goal in mind that this architecture will fulfil the requirements of the whole system. This process can be repeated recursively. The overall task described by the system requirement is decomposed into sub-tasks annotated to the components (e.g. "Switch off power-stage when motor current gets higher than x within y ms [ASIL D]" → C_1: "Measure motor current"; C_2: "Compare Current measurement to x and trigger switchoff command if higher"; C_3: "Switch off power-stage on receiving switch-off command".) At each port, signal properties like accuracy, delay or safety integrity level are specified in an assertion language that serves for both assumptions and guarantees. A component that consumes a signal at its input can rely on the specified property as an assumption. A component that provides some signal at its output must guarantee the specified property so that this guarantee becomes a requirement for this component. The architect has to negotiate acceptable assumptions and guarantees with component suppliers, or look for readily available components that fulfil the required guarantees under the given assumptions. Where input/output ports of components are directly connected to input/output components of the surrounding system, assumptions and guarantees propagate one level up. Verification is done by checking if every assumption is satisfied by a guarantee (referring to a connected pair of output port and input port) that is at least as strong as supposed by the assumption. An example of the usage of Interface Contacts is shown in Figure 1:

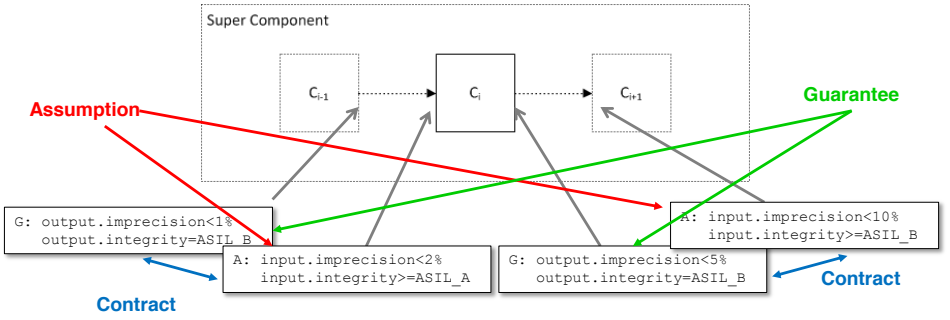


Figure 1: An example using Interface Contracts

The component under consideration is C_i. It consumes an input value from its predecessor in the signal chain, C_{i-1}, and provides an output signal to its successor C_{i+1}. It assumes certain quality properties regarding the input signal, e.g. an imprecision (with respect to the true value in the real world) of less than 2% and a safety integrity of at least ASIL A (according ISO 26262 scale). The producer C_{i-1} of the signal guarantees an imprecision of less than 1% for the same signal, which is better than assumed, and integrity of ASIL B. Therefore the contract between these two components is valid and the components are compatible to each other.

3.1 Syntax of the proposed Contract Annotation Language

To automate the described concept by a software tool, it is necessary to define a syntax (definition of well-formedness off the semi-formal notation) for assertions (assumptions or guarantees), even if the semantics of terms like “imprecision” or “ASIL” is not defined and must be interpreted by humans. The notation should be acceptably understandable to domain engineers. For first experiments, we invented a contract annotation language that has a high degree of similarity with well-known OO programming languages such as Java or C++, and that allows expressing three types of relevant properties: delay, accuracy and automotive safety integrity level (ASIL). An example of the syntax structure is given below:

```
componentName.portName.propertyName relational_operator value [unit]
```

Listing 1: Syntax example of the proposed contract notation language

The assumptions and guarantees are mapped to elements of the modeled system by their unique names. The *componentName* maps them to a unique component of the system, while the *portName* maps the given element to a precise signal port (input or output) of that component. Signals are logically typed as event signals, enumerative signals or continuous signals. These can be compared to reference values, either as binary comparison or by comparing their relation to a given value (equal, greater, or less than the value), which is a precondition for dominance checking. The annotation language allows some predefined ordinal scale values such as ASIL_A to ASIL_D and the optional use of a set of relative or physical units for continuous values as %, *ms* or *V* in order to prevent user errors by confusion. The same syntax applies to assumptions and guarantees; the actual role of an element is deducted by its placement. The prototypical tool provides syntax checking for the assumptions and guarantees typed in by the user.

3.2 Proposed Workflow for computer-aided checking

The proposed workflow for the computer-aided checking of the contracts, as

implemented in the prototypical tool in Section 4, is as follows: (i) the system is graphically modeled using components and their connecting signals, (ii) components are annotated with their assumptions and guarantees directly in the model using the proposed annotation language, (iii) the analysis algorithm is run, doing an automated check for completeness (there are no assumptions without matching guarantees) and consistency (each guarantee implies the assumption it is linked to) of the contracts, (iv) the results are visualized and can be inspected. Following the workflow of the development processes of the embedded system, this sequence can be repeatedly carried out in an iterative fashion.

3.3 Contract Checking Algorithm

To perform the mentioned type of contract checking, an algorithm has been designed and implemented in the prototypical tool. The concept of this algorithm works as follows. First, all components are collected and saved in a list. All children of a component are added to the list of components recursively. For all assumptions, corresponding guarantees are searched, which are selected by the fact that they refer to the same signal (i.e. the one that connects output of one component to input of another) and property (i.e. imprecision or time delay or safety integrity level). If a guarantee is found, it is entered into the same contract dataset where the assumption came from. If no suitable guarantee is found, the assumption is checked for transferred assumptions, i.e. assumptions that have their origin on the next higher level of hierarchy. In case a guarantee was found and the units of guarantee and assumptions are equal, the values are compared with the relations defined in the guarantee and assumption. If the guarantee provides a better or equal signal quality than claimed by the assumption (e.g. assumption: Safety Integrity Level of ASIL A, guarantee: Safety Integrity Level of ASIL B), the contract is marked with an OK label, otherwise it is marked as NOK (not ok). If no guarantee is found, the contract is set to MISSING, which indicates, that no appropriate guarantee for the assumption was found. Additionally the guarantees are also checked for transferred guarantee compliance. To solve the problem of the verification of transferred statements, assumptions and guarantees have the same super-class so that both can be used in the position of an assumption or a guarantee in the contract.

4 Prototypical Tool

For evaluation of the idea a prototypical tool was created. The demonstrator includes the necessary functionality for modeling a system in a block diagram (actually a simplified form of a SysML Internal Block Diagram), annotating assumptions and guarantees for a set of relevant properties, automatically matching and checking contracts using the proposed algorithm and visualizing the results in a table. The user interface is based on the widespread *Eclipse* platform with its *Rich Client Platform* extension facility. The graphical editor is based on the *Graphiti* framework, which allows a quick prototyping of almost any kind of graphical representation based on a meta-model of the items to be modelled using the *Eclipse Modeling Framework* (EMF) [St08]. The proposed user

interface is based on two main views, illustrated in Figure 2: (i) the editor view showing the graphical representation of the system under construction and (ii) a table based view showing the results after analyzing the system contracts. In addition, it contains some supporting views and palettes for quicker access to functions and projects.

The editor displays a graphical model of the static architecture of the system under development (SysML Internal Block Diagram). Rectangles denote components, small squares with right or left pointing arrows on the outer borders of a component denote output and input ports, respectively, and lines connecting these ports denote connections of the ports via signals. The hierarchical nature of the modeled system is denoted by components that graphically contain other components. The model elements can be added by dragging them from a supporting palette view. In a later professional version, the model would be importable from existing UML/SysML modeling tools. Also, an import of block models from Simulink would be desirable; we have already implemented this feature in another related master thesis [Ra14].

After creating a system model, the system architect annotates assumptions and guarantees. These are added using the annotation language defined in Section 3.1. The annotated assumptions and guarantees are visualized as elements inside a component.

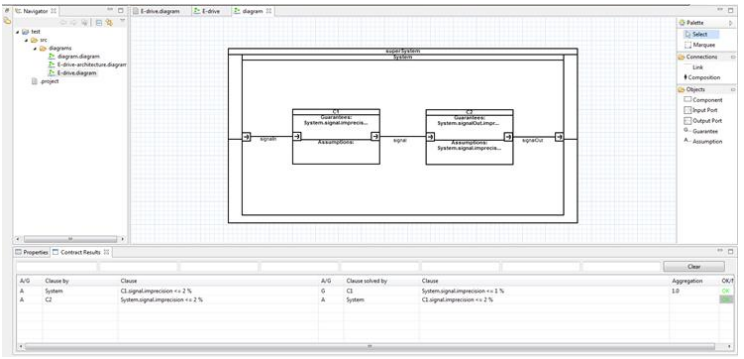


Figure 2: Screenshot of the tool GUI

The contract check engine implements the algorithm explained in Section 3.3 and can manually be started from the tool GUI. The results are then displayed in the tabular *Result View*. The table lists each detected contract with the following details in distinct cells of the table: The component that declared the assumption and the component that declared the guarantee, the guarantee (G) and the assumption (A) itself in the syntax explained above, and its type (G or A). A field "Aggregation" shows the result of the calculation of a guarantee and its referenced assumption by using the declared operation. A last column shows the verification results (OK, NOK and MISSING). By colors, the table clearly marks the remaining issues (contracts that do not match and assumptions with missing guarantees). Thereby, the correctness of the intended system architecture or the aptitudes of a reused component for its new context become visible at a glance.

5 An Automotive Electric Drive Case Study

The method and the tool have been evaluated based on the example of a simplified automotive electric traction drive system, as first introduced in [Ka12]. The system consists of a three-phase synchronous machine (PSM) and an inverter with 6 IGBT transistors as power switches, along with the required driver circuitry. The power stage is commanded by PWM (pulse width modulation) signals generated by a microcontroller, on which a complex software system is running. Among many other tasks, the software performs the closed-loop control algorithm (field-oriented control) to control the torque of the PSM by modulating a voltage system. The torque reference value is delivered via a data bus and represents the value derived from the accelerator pedal angle. The actual values of a number of physical quantities required for control (e.g. phase currents or rotor angle) are measured by suitable sensors with corresponding analog signal conditioning circuitry and then converted to digital numbers. This example is sufficiently similar to real vehicle drive systems, and for such systems, a set of safety goals is commonly known (we have meanwhile implemented an operating electric drive controller for a 1:8 scale toy car on this base, and have started developing a technical safety concept for it [Fe14], which is inspired from actual industry projects). A typical example of a safety goal is to prevent the machine from applying unauthorized or excessive torque, which could lead to an uncontrollable situation and in consequence to a road accident. Breaking down this requirement stepwise from vehicle level to controller level using our approach, one resulting technical safety requirements could, for instance, read as “It shall be detected if a phase current is higher than the commanded current by more than 10A, and if so, the power stage shall be disabled within at maximum 200 ms. [ASIL D]”. During system refinement and many iterations involving different safety analyses (such as FMEA or Fault Tree Analysis), safety mechanisms (e.g. signal plausibility checks) are added to the system, and the initial safety goals are transformed to more specific requirements, e.g. to detect that a phase current measurement value is not trustworthy, and to do so within a given time and with a defined accuracy. The example presented here starts at that point and deals with some of those checks.

For applying our semi-formal contract approach to this system, the first step is to model its static architecture. Each block and each signal is annotated with a verbal description of what it does or means (e.g. “compares phase current values and generates shut-off command when they seem no longer to be in a plausible correlation”). Figure 3 illustrates the editor view containing the static architecture of the drive model.

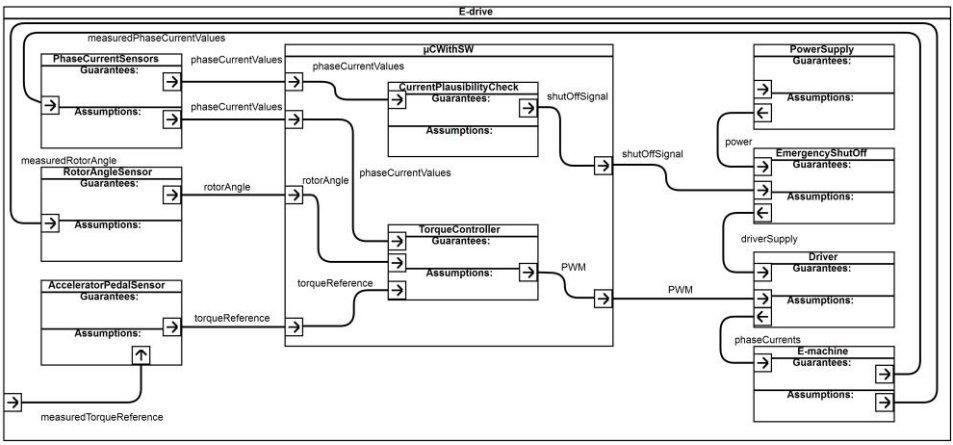


Figure 3: Static architecture of the electric drive system, containing some safety mechanisms

Then, the requirements are broken down to the components of the system and coded into assumptions and guarantees. Taking the phase current sensor as an example, typical requirements for these are (i) the delay of the (analog) output value w.r.t. the true value, (ii) its accuracy or imprecision w.r.t the true value, and their (iii) safety integrity level. These properties are typical examples of properties that are today still difficult to prove formally, but which can be judged by experts (e.g. using timed tolerance bands, evaluating step responses, etc.). For each signal, the architect proposes an assertion. For the component that produces the signal at its output, it becomes a requirement to *guarantee* the denoted signal properties, and the component that consumes the signal at its input may be designed under the *assumption* that the signal fulfills the denoted properties. Thereby, the architect creates a contract between producer and consumer component. How to distribute e.g. the allowed overall delay of a safety reaction onto the components is left to his free decision, based on his experience and on negotiation.

In our example, the producer component is the set of current sensors with its related analog signal conditioning ("PhaseCurrentSensors"), the signal is the measurement value at the input of the microcontroller and the consumer component is the microcontroller ("μCWithSW"). Listing 2 shows these refined requirements modeled as assumptions.

```
Assumptions for the μCWithSW on the PhaseCurrentSensor:
μCWithSW.phaseCurrentValues.delay <= 20 ms
μCWithSW.phaseCurrentValues.imprecision <= 2mA
μCWithSW.phaseCurrentValues.ASIL >= B
```

Listing 2: Assumptions for the phase current sensor

The guarantees of a candidate component for "PhaseCurrentSensors" can be stated by the system architect, e.g. based on the datasheets. They are provided in Listing 3.

```
Guarantees given by the PhaseCurrentSensors:
phaseCurrentSensors.phaseCurrentValues.delay <= 15 ms
phaseCurrentSensors.phaseCurrentValues.imprecision <= 1mA
phaseCurrentSensors.phaseCurrentValues.ASIL >= B
```

Listing 3: Guarantees for the phase current sensor

Once the system is modeled and annotated with assumptions and guarantees, it is automatically analyzed by the analysis engine. Figure 4 illustrates the *Results View* after the analysis engine was run. For the above example the result table lists three rows, one for each contract. In this case all contracts are evaluated to 'OK'.

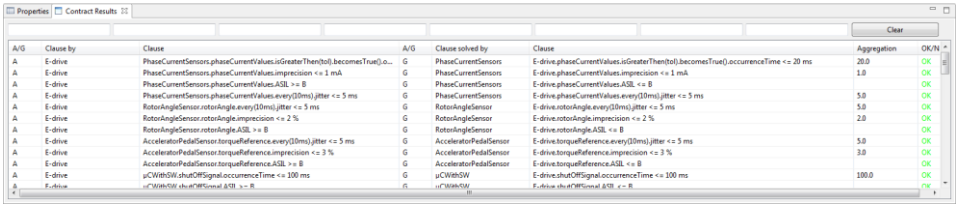


Figure 4: Result view after analyzing the system

The analysis traverses the signal connections derived from the static architecture and checks that all assumptions and guarantees are syntactically correct, that for each assumption a corresponding guarantee is found and that the assumption implies the guarantees. It is automatically evaluated, that the delay property guaranteed by PhaseCurrentSensors is sufficiently good for the needs of µCWithSW, because 15 ms is better than the required 20 ms, and a semantics for “better than” regarding delays (in this case a lower value is better) has been implemented in the tool. In contrast, due to the lack of full formal semantics, it cannot be checked automatically if the sensor is actually able to deliver the signal within 15 ms and if the microcontroller is actually be able to perform some safety reaction on time whenever it gets the current values within a delay of 20 ms. This assessment is still left to the expert’s manual evaluation, e.g. by calculations, simulations, or by integration testing.

For evaluating the usability of the method and the tool, we have presented the tool and the case study to several safety consultants of our company: they found the considerations on the electric drive valid, and after initial instructions, they were all able to specify some properties on their own and to interpret the check results from the tool.

6 Conclusion and Further Research Agenda

In this paper, we have proposed a procedure to apply Interface Contracts to the task of requirements decomposition and allocation onto components in the domain of safety-critical automotive systems. We provided an automated method to check the consistency

of the system architecture and we have developed a prototypical tool and applied all this to an electric vehicle drive as an example. The approach is promising, as we obtained positive feedback from several safety engineers who create safety concepts for automotive systems of numerous automotive OEMs and suppliers in their daily work. The ease of use is due to the graphical user interface with familiar component representation and the annotation language that is similar to well-known OO programming languages and allows the usage of physical units. In the future, the static architecture could be imported from existing UML/SysML design tools or Simulink, and signals could be imported from a data dictionary usually available.

We plan to extend the approach to further properties of port signals, but also to contracts that do not apply to ports, but to the component as a whole (component contracts). An example could be the assurance of sufficient processing resources for a software component.

Further it seems promising to combine this contribution with other recent development activities of ours: in a bachelor thesis, we have been performing experiments with parametrized templates for requirements, allowing direct references to signal names from a signal dictionary [Kü10]. Some of these templates appear to be promising candidates for requirements to the top-level systems that can be automatically translated into our annotation language. Secondly, we developed a tool that is able to extract the component hierarchy and the signal connections from Simulink tools and to create Component Fault Tree frames out of it [Ra14]. We plan to integrate these building bricks into a comprehensive Eclipse-based tool-chain for system and safety development, allowing the specification of contracts for the normal system function, the analysis of failures defined in terms of violations of these contracts, and the addition of safety contracts that describe the mitigation of the effects of these failures by additional safety mechanisms.

References

[Be12] Benveniste, A., Caillaud, B., Nickovic, D., Passerone, R., Raclet, J., Reinkemeier, P., Sangiovanni-Vincentelli, A., Damm, W., Henzinger, T., Larsen, K. 2012. "*Contracts for Systems Design"*. INRIA, Tech. rep., 2012.

[BHM03] Bate, I., Hawkins, R. and McDermid, J. 2003. "A Contract-based Approach to Designing Safe Systems". In Proceedings of the 8th Australian Workshop on Safety Critical Systems and Software - Volume 33, Darlinghurst, Australia, 2003, S. 25–36

[BKW15] Battram, P, Kaiser, B. and Weber, R.: A Modular Safety Assurance Method considering Multi-Aspect Contracts during CPS Design. RESACS 2015

[Ba10] Baumgart, A., Reinkemeier, P., Rettberg, A., Stierand, I., Thaden, E. and Weber, R. 2010. „A Model-Based Design Methodology with Contracts to Enhance the Development Process of Safety-Critical Systems". In SEUS, Springer, 2010; S. 59-70.

[Bo11] Boström, P. “Contract-Based Verification of Simulink Models”. ICFEM, 2011

[Fe07] Fenn, J., Hawkins, R., Williams, P. and Kelly, T. 2007. “Safety case composition using contracts-refinements based on feedback from an industrial case study”. The Safety of Systems. Springer, 2007; S. 133–146.

[Fe14] Feyerherd, F.: “Methodische Erstellung eines Sicherheitskonzeptes für einen elektrischen Antrieb“. Master thesis, HTW Berlin, 2014

[Ka12] Kaiser, B. 2012. “Functional Safety in Component-Oriented Development Contexts”. Philotech Symposium "Achieving Functional Safety in Avionics and Automotive", Oct. 17, 2012, Berlin

[Kü10] Kühl, C. 2010. “Formalisierung von Requirements durch Nutzung von Templates”. Master thesis, Freie Universität Berlin, 2010.

[Me92] Meyer, B. 1992. “Applying ’design by contract“. Computer (10)IEEE, 1992, S. 40–51.

[Ra14] Ramich, V. 2014. “Teilautomatische Erstellung von Component-Fault-Trees aus Simulink-Modellen”. Master thesis, Universität Kassel, 2014.

[Sc14] Schreiter, L., Bresolin, D., Capiluppi, M., Raczkowsky, J., Fiorini, P. and Woern, H. 2014. “Application of contract-based verification techniques for hybrid automata to surgical robotic systems”. In European Control Conference (ECC), 2014

[So13] Sonski, S.: Contract-based modeling of component properties for safety-critical systems. Master Thesis, Hochschule Darmstadt, 2013

Acknowledgement: This research has been partly supported by the BMBF (German Federal Ministry of Education and Research) funded research project SPES_XT.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Automatische Verlinkung von Anforderungen und Testfällen - Fallbasierte Filterung

Thomas Noack[1] Steffen Helke[2]

Abstract: Der systematische Einsatz von Techniken zur Wiederverwendung von Entwicklungsartefakten führt nachweislich zu einer drastischen Kostenreduktion in Softwareprojekten. Das Potential zur Wiederverwendung ist speziell im Automotivebereich besonders hoch, da sich die Anforderungen von Baureihe zu Baureihe nur teilweise ändern. Konzepte zur Wiederverwendung sind effektiv, wenn sie durch die eingesetzten Entwicklungswerkzeuge unterstützt und automatisiert werden. Deshalb haben wir in jüngeren Arbeiten eine Technik zur automatischen Verlinkung von Testfällen mit wiederverwendeten Anforderungen vorgestellt, die als Erweiterung von DOORS umgesetzt ist. Die Technik greift auf wiederverwendete Anforderungen zurück, die aus einer Vorgängerbaureihe übernommen wurden und stellt durch automatisch erzeugte Links sicher, dass diese durch die selben Testfälle abgesichert werden, wie die Anforderungen der Vorgängerbaureihe. Suspekte Verlinkungssituationen, die z.B. durch die Modifikation wiederverwendeter Anforderungen entstehen können, werden dabei erkannt und mit einem allgemeinen Prüfhinweis versehen. Dieses Papier erweitert dieses Verfahren um eine fallbasierte Filterung mit dem Ziel, suspekte Verlinkungssituationen detaillierter zu beschreiben und konstruktive Vorschläge zu deren Auflösung bereitzustellen.

Keywords: Anforderungen, Testfälle, Wiederverwendung, Traceability, Case-based Reasoning

1 Motivation

Die Wiederverwendung von Entwicklungsartefakten hilft entscheidend, die Kosten zur Erstellung von Software zu senken. Speziell im Automotivebereich ist dieses Vorgehen zielführend, da sich die Anforderungen von Baureihe zu Baureihe nur teilweise ändern. Zugleich birgt Wiederverwendung aber auch Gefahren, denn jedes wiederzuverwendende Artefakt muss zunächst sorgfältig geprüft werden, ob es passgenau ist und ob es adäquat im neuen Kontext verwendet werden kann. Eine sorgfältige Prüfung ist zeitaufwendig, fehleranfällig und verursacht wiederum hohe Kosten, insbesondere dann, wenn sie manuell durchgeführt werden muss. Deshalb wird intensiv daran geforscht, wie prozessunterstützende Entwicklungswerkzeuge durch Komponenten zur automatisierten Wiederverwendung ergänzt werden können. In jüngeren Arbeiten haben wir eine Technik zur automatischen Verlinkung von Testfällen mit wiederverwendeten Anforderungen vorgestellt, die in dem Anforderungsmanagementwerkzeug DOORS umgesetzt ist [NHK14]. Das Verfahren setzt voraus, dass für wiederverwendete Anforderungen nachvollzogen werden kann, aus welchen Anforderungen der Vorgängerbaureihe sie abgeleitet wurden. Ferner wird davon ausgegangen, dass die Anforderungen der Vorgängerbaureihe mit Testfällen

[1] Daimler Center for Automotive IT Innovations — TU Berlin, thomas.noack@tu-berlin.de
[2] Brandenburgische Technische Universität Cottbus-Senftenberg, steffen.helke@b-tu.de

verlinkt sind, die diese absichern. Basierend auf den Informationen werden Links zwischen Testfällen und wiederverwendeten Anforderungen automatisch erzeugt. Das Verfahren kann bereits mit dem Wegfall von Anforderugen oder Änderungen der wiederverwendeten Anforderungen umgehen. In so einem Fall wird eine suspekte Verlinkungssituation detektiert und ein allgemeiner Prüfhinweis eingefügt. Eine präzise Beschreibung der Verlinkungssituation war bisher allerdings noch nicht möglich, da (1) keine klassifizierenden Eigenschaften für die wiederzuverwendenden Artefakte vorlagen und (2) kein geeignetes Suchverfahren existierte, das auf kritische Wiederverwendungssituationen hinweisen konnte. Dieses Papier stellt eine Technik vor, die das bisherige Verfahren um eine fallbasierte Filterung erweitert, so dass auf Grundlage von Erfahrungswissen suspekte Verlinkungssituationen genauer charakterisiert und konstruktive Vorschläge zu deren Auflösung getätigt werden können.

Das Papier ist folgendermaßen aufgebaut. In Abschnitt 2 werden grundlegende Begriffe eingeführt und die Arbeit in einen größeren Zusammenhang gestellt. Abschnitt 3 führt in das sogenannte WvT-Problem ein, das präzise die Fragestellung beschreibt, ob ein Testfall mit einer wiederverwendeten Anforderungen verlinkt werden sollte. Ferner gehen wir in diesem Abschnitt auf verwandte Arbeiten ein. In Abschnitt 4 stellen wir das Konzept der Fallbasierten Filterung vor. Wir beschreiben, wie Artefakte genauer klassifiziert werden können und wie mit Hilfe von Ähnlichkeitsmaßen eine Filterung erfolgen kann. In Abschnitt 5 stellen wir schließlich vier Aktivitäten vor, die zur methodischen Umsetzung des Verfahrens notwendig sind. In Abschnitt 6 fassen wir die Ergebnisse zusammen und geben einen Ausblick auf weiterführende Arbeiten.

2 Grundlagen

Die Spezifikation von Anforderungen und Testfällen ist eine komplizierte Tätigkeit, die im Projekt nicht einfach nebenbei erledigt werden kann. Eine besondere Bedeutung wird ihnen zuteil, wenn Anforderungen von Herstellern an Lieferanten übergeben werden und diese Lieferanten die bestellten Systeme realisieren. Um nachweisen zu können, dass die bestellten Systeme korrekt umgesetzt wurden, werden herstellerseitig Testfälle spezifiziert. Diese Testfälle sind mit den Anforderungen verlinkt. Die Verlinkung hat mehrere Vorteile, sie lokalisiert nicht korrekt umgesetzte Anforderungen und ermöglicht die Bewertung der Testabdeckung von Anforderungen.

2.1 Anforderungsentwicklung

Anforderungen entstehen nicht von jetzt auf gleich - im Gegenteil, sie resultieren aus einem systematischen Anforderungsentwicklungsprozess. Dieser Prozess wird in drei Kernaktivitäten Gewinnung, Übereinstimmung und Dokumentation sowie zwei Querschnittsaktivitäten Validierung und Management gegliedert. Die einzelnen Aktivitäten sind in der Standardliteratur ausführlich beschrieben [Po08]. Die bereits früher vorgestellte Technik zur automatischen Verlinkung von Testfällen mit wiederverwendeten Anforderungen

[NHK14] ordnet sich in die Querschnittsaktivität *Management von Anforderungen* ein. Hier werden drei Bereiche unterschieden: *Priorisierung* von Anforderungen, *Verwaltung der Änderung* von Anforderungen sowie die *Verfolgbarkeit* von Anforderungen.

Der Kern unserer Methode ist die Verfolgbarkeit der Wiederverwendung von Anforderungen hinsichtlich des Tests. Wiederverwendung ist hierbei als Änderung zu verstehen, wenn wiederverwendete Anforderungen modifiziert werden. Damit ordnet sich unsere Methode in den Bereich der Anforderungsverwaltung (engl.: requirements management) ein.

2.2 Testentwicklung

Wie im Anforderungsentwicklungsprozess entstehen Testfälle nicht von jetzt auf gleich. Auch Testfälle sind das Resultat eines systematischen Entstehungsprozesses. Der fundamentale Testprozess wird in die fünf Arbeitsschritte Testplanung und Teststeuerung, Testanalyse und Testdesign, Testrealisierung und Testdurchführung, Testauswertung und Bericht sowie Abschluss der Testaktivitäten gegliedert [SL05]. Die automatische Verlinkung von Testfällen mit wiederverwendeten Anforderungen ordnet sich in den Arbeitsschritt *Testplanung und Teststeuerung* ein: In der *Testplanung* werden die Umfänge des Testens definiert, die während der *Teststeuerung* verwaltet werden. Das Ergebnis der Testplanung ist das Testkonzept. In diesem wird beschrieben, auf welcher Testumgebung (Wo?) während welcher Teststufe (Wann?) welche Testobjekte (Was?) abzusichern sind und welche Testziele (Wozu?) dabei verfolgt werden. Das Testkonzept legt somit fest, welche Testfälle mit welchen Anforderungen verlinkt sein müssen. In jüngeren Arbeiten haben wir beschrieben, wie mit Hilfe des Testkonzepts die Testabdeckung von Anforderungen ermittelt werden kann [No13]. In der Teststeuerung wird überwacht, ob das Testkonzept erfüllt ist und in welchem Umfang Abweichungen auftreten.

Die in dieser Arbeit vorgestellte Methode verlinkt Anforderungen mit wiederverwendeten Anforderungen. Das Ziel dieser Verlinkung ist es, die Nachvollziehbarkeit in der *Testplanung und Teststeuerung* zu fördern, wenn Testfälle auf der Grundlage wiederverwendeter Anforderungen wiederholt werden sollen.

2.3 Verfolgbarkeit

Im Deutschen wird der englische Begriff Traceability unterschiedlich übersetzt. Während einige Autoren den Begriff Nachvollziehbarkeit verwenden, greifen andere auf den Begriff Zurückführbarkeit zurück. Weitere, häufig verwendete Begriffe sind Rückverfolgbarkeit, Nachverfolgbarkeit oder Durchgängigkeit. Zuletzt erschien ein englischsprachiges Buch, in dem die Traceability-Begriffswelt umfassend definiert wird [CHGZ12]. Basierend auf diesem Buch übersetzen wir Traceability mit Verfolgbarkeit [No15]. Verfolgbarkeit ist die generelle Fähigkeit, Artefakte miteinander zu verlinken und diese Verlinkung zu verfolgen. Nachvollziehbarkeit (engl.: comprehensibility) ist als Zweck der Verfolgbarkeit zu verstehen. Weil wir Anforderungen und Testfälle verfolgen, ordnet sich unsere Arbeit in den Bereich der Anforderungs- und Testverfolgbarkeit ein. Dabei ist es das Ziel, Wiederverwendung von Anforderungen und Wiederholung von Testfällen nachzuvollziehen.

3 Anforderungs- und Testverfolgbarkeit

Anforderungen und Testfälle können nicht getrennt voneinander betrachtet werden. Sie stehen in einem Verhältnis, da Testfälle die korrekte Umsetzung von Anforderungen absichern. Dabei muss sichergestellt sein, dass die richtigen Testfälle mit den richtigen Anforderungen verlinkt sind. In diesem Abschnitt wird das WvT-Problem vorgestellt. Es entsteht, wenn Anforderungen wiederverwendet werden und entschieden werden muss, ob die existierenden Testfälle noch dazu geeignet sind, die wiederverwendeten Anforderungen abzusichern.

3.1 Anforderungen und Testfälle in der Praxis

Während der Anforderungs- und Testentwicklung entstehen zahlreiche Anforderungen und Testfälle für eine Vielzahl von softwarelastigen Systemen. Die Anforderungen werden dabei in Systemlastenheften (SLH) und die Testfälle in Systemtestspezifikationen (STS) gepflegt. Um Nachvollziehbarkeit zu ermöglichen, sind die Anforderungen und Testfälle verlinkt. Bild 1 zeigt einen Ausschnitt aus Spezifikationsdokumenten. Die obere linke Box zeigt ein Systemlastenheft einer Vorgängerbaureihe. Die obere rechte Box zeigt ein Systemlastenheft der aktuellen Baureihe. Dieses Systemlastenheft wurde aus der Vorgängerbaureihe wiederverwendet, indem es kopiert und angepasst wurde: In der Anforderung *ziel-134* wurde die Spezifikation der *Waschunterbrechung* entfernt. Die untere Box zeigt eine Testspezifikation. Die Testfälle *93* und *94* sichern eben diese *Waschunterbrechung* von Anforderung *134* ab. Das Beispiel deutet die Problemstellung dieser Arbeit an: Sind die Testfälle *93* und *94* dazu geeignet, die Anforderung *ziel-134* abzusichern?

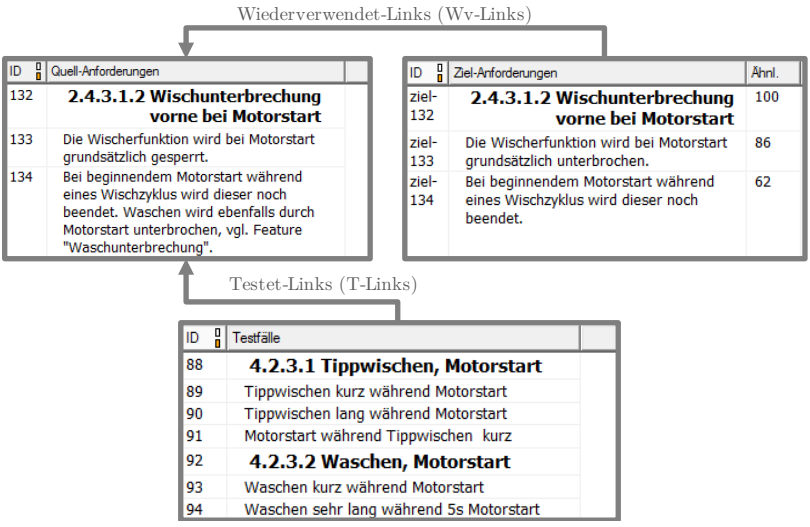


Abb. 1: Beispiele für Systemanforderungen und Systemtestfälle

3.2 WvT-Problem

Die Methode zur Verlinkung von Testfällen und Anforderungen basiert auf einem spezifischen Verlinkungsmodell, das in der industriellen Praxis beobachtet wurde. Bild 2 zeigt dieses Modell, das als Wiederverwendet-Testet- oder WvT-Problem bezeichnet wird.

Wv: Wiederverwendet. Im Rahmen eines neuen Baureihenprojekts werden SLH kopiert und an das neue Umfeld angepasst. Somit entstehen zwei SLH, das SLH_{Quell} der Quell-Baureihe und das SLH_{Ziel} der Ziel-Baureihe. Diese stehen in einem Wiederverwendungsverhältnis: SLH_{Ziel} wiederverwendet SLH_{Quell}.

T: Testet. Die STS wird nicht kopiert und angepasst. Stattdessen werden die Testfälle mit den neuen Anforderungen des SLH_{Ziel} verlinkt. Vor dieser Verlinkung befinden sich die STS und das SLH_{Quell} in einem Testverhältnis: STS testet SLH_{Quell}

WvT-Problem und WvT-Verlinkung. In Bild 2 zeigt ein Wv-Link von einer Ziel-Anforderung zu einer Quell-Anforderung. Weiterhin zeigt ein T-Link von einem Testfall zu einer Quell-Anforderung. Im WvT-Problem existiert der gestrichelt dargestellte T-Link vom Testfall zur Ziel-Anforderung noch nicht. Damit beschreibt das WvT-Problem die Situation, bevor der Testfall ein Testverhältnis mit der Ziel-Anforderung eingeht. Die WvT-Verlinkung setzt den fehlenden T-Link auf der Grundlage des Wv-Links. Dabei werden suspekte Verlinkungssituationen, sogenannte WvT-Inkonsistenzen, aufgedeckt [NHK14].

Das Setzen des fehlenden T-Links wird als *Lösen des WvT-Problems* bezeichnet. Der aktuelle Stand der Technik kennt zwei Vorgehensweise zur Lösung von WvT-Problemen: die manuelle Verlinkung und die Verlinkung nach dem Anlegen der SLH-Kopie. Moderne Fahrzeug haben bis zu 100 Steuergeräte und somit mindestens 100 Systeme. Wenn pro System ein SLH und eine STS mit jeweils durchschnittlich 5.000 Anforderungen und Testfällen existieren, führt das erste Vorgehen offensichtlich zu einer Zeit- und Genauigkeitsproblematik. Im zweiten Vorgehen findet die Verlinkung zum falschen Zeitpunkt statt: Nach dem Kopieren aber vor dem Anpassen des SLH_{Ziel}. Ungeachtet der Vorgehensweise kann ein Umstand häufig beobachtet werden: Von Baureihe zu Baureihe existieren immer weniger T-Links. Das bedeutet in der Praxis zwar nicht, dass weniger getestet wird - wohl aber, dass der Test schlechter nachvollziehbar ist.

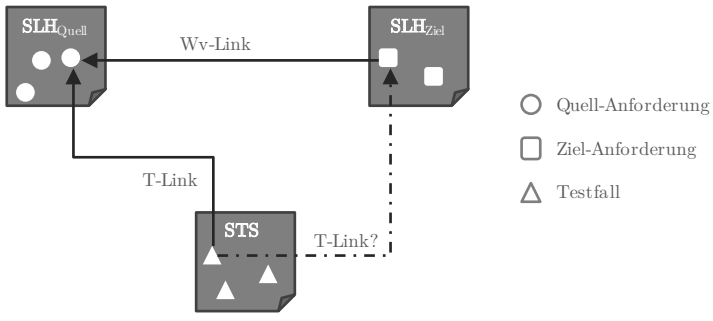


Abb. 2: WvT-Problem und WvT-Verlinkung

3.3 Verwandte Arbeiten

Das WvT-Problem ist ein Verlinkungsmodell, das das Wiederverwendungsverhältnis zwischen Anforderungen beschreibt. Es ordnet sich somit in den Bereich der Anforderungsverfolgbarkeit (engl.: requirements traceability) ein. Neben der Anforderungswiederverwendung spiegelt das WvT-Problem auch das Testverhältnis zwischen Anforderungen und Testfällen wider. Damit ordnet es sich auch in den Bereich der Testverfolgbarkeit (engl.: test traceability) ein. Die WvT-Verlinkung und die damit verbundenen Analysen stellen Verlinkungstechniken dar.

Während Verlinkungsmodelle (engl.: traceability information models) die beteiligten Artefakte und die möglichen Links zwischen ihnen benennen, beschreiben Verlinkungstechniken den Umgang mit den Modellen. Hier existieren bereits zahlreiche Arbeiten, von denen im Folgenden die wichtigsten vorgestellt werden.

3.3.1 Verlinkungsmodelle

In den 1990er Jahren galt der Trend dem Hypertext. Die Idee war es, Anforderungen in Hypertext zu formulieren und durch Hyperlinks miteinander zu verlinken [PH95]. Da Hyperlinks generisch sind, war es in frühen Verlinkungsmodellen möglich, grundsätzlich alles mit allem zu verlinken. Obwohl es möglich ist, in den generellen Modellen Links von/zu Testfällen zu definieren, fehlen den vorgestellten Modellen explizite T-Links.

Neuere Arbeiten beschreiben Verlinkungmodelle mit expliziten T-Links [Ki08][AI11]. Diese Modelle streben jedoch die Beschreibung eines möglichst vollständigen Verlinkungsmodells an, um den gesamten Softwarelebenszyklus abzubilden.

Das WvT-Problem grenzt sich davon ab, indem es sich explizit auf einen einfachen Sachverhalt beschränkt: Die Wiederverwendung von Anforderungen und die damit einhergehende Wiederverwendung von T-Links. Damit ist das WvT-Problem kein generelles, sondern ein spezialisiertes Verlinkungsmodell. Zusätzlich führt das WvT-Problem den noch nicht verbreiteten Wv-Link zwischen wiederverwendeten Anforderungen ein.

3.3.2 Verlinkungstechniken

Es lassen sich zwei Arten von Verlinkungstechniken unterscheiden: (1) Die automatische Verlinkung und (2) die automatische Linkanalyse. Im Forschungsstand existieren Ansätze zur ereignisbasierten (EBV), regelbasierten (RBV), variabilitätmodellbasierten (VBV), kostenbasierten (KBV), szenariobasierten (SBV), zielbasierten (ZBV) oder information-retrieval-basierten (IRV) Verlinkung. Diese Verlinkungstechniken wurden bereits in verschiedenen Publikationen beschrieben [RWA07][To12]. Die in diesem Papier vorgestellte Methode erweitert den Forschungsstand um die wiederverwendungsbasierte Verlinkung (WBV). Zudem werden einige der Techniken aufgegriffen, um die WvT-Verlinkung zu erweitern: Die fallbasierte Filterung ist eine IRV.

4 Fallbasierte Filterung

Die WvT-Verlinkung verlinkt Testfälle mit wiederverwendeten Anforderungen gemäß Abbildung 2 auf der Grundlage von bestehenden Links zwischen den beteiligten Artefakten. Artefakte haben aber nicht nur Links. Sie besitzen auch Eigenschaften, die sie klassifizieren. Im Folgenden werden Techniken beschrieben, wie diese Eigenschaften während der WvT-Verlinkung berücksichtigt werden können. Damit wird es möglich, detaillierte Prüfhinweise bereitzustellen, die den Grund suspekter Verlinkungssituationen aufzeigen.

4.1 WvT-Fall

Abbildung 3 zeigt einen angereicherten WvT-Fall. Ein angereicherter WvT-Fall unterscheidet sich lediglich durch einen Prüfhinweis vom angereicherten WvT-Problem. Das Ziel der fallbasierten Filterung ist es, für aktuelle angereicherte WvT-Probleme ähnliche WvT-Fälle in der Falldatenbank zu suchen und den Prüfhinweis in dem aktuellen WvT-Problem zu hinterlegen. Fälle entsprechen demnach früher gelösten Problemen.

Ein WvT-Fall bzw. WvT-Problem wird angereichert, indem klassifizierende Eigenschaften (KE) der Quell- und Ziel-Anforderungen sowie des Testfalls gespeichert werden. In der Praxis entsprechen KE DOORS-Attributen. Im Beispiel besitzen die beiden Anforderungen Schnittstellen. Während die Quell-Anforderung eine Schnittstelle zum Zündschloss hat, besitzt die Ziel-Anforderungen eine weitere neue Schnittstelle zum Motorsteuergerät. Die Schnittstellen haben sich also geändert. Der beteiligte Testfall ist auch klassifiziert. Er ist ein Schnittstellentestfall. Dieses WvT-Problem ist suspekt, weil sich die Schnittstellen der Anforderungen geändert haben und ein Schnittstellentestfall beteiligt ist.

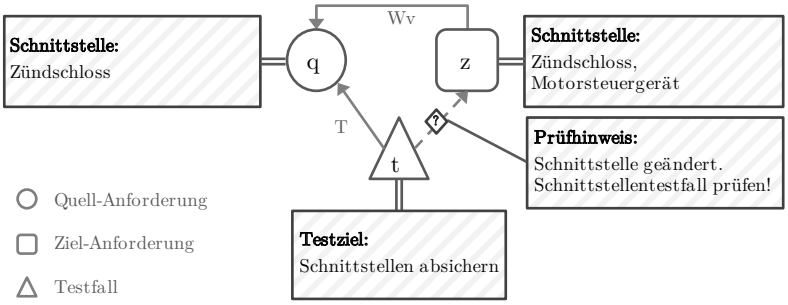


Abb. 3: WvT-Fall mit klassifizierenden Eigenschaften

Um zu erkennen, ob ein aktuelles WvT-Problem einem WvT-Fall aus der Falldatenbank ähnelt, werden binäre Ähnlichkeitsmaße [Br13] eingesetzt. Diese ermitteln, wie viele gleiche bzw. unterschiedliche Elemente die Ausprägungsmengen einer KE besitzen und quantifizieren die Ähnlichkeit. In Abbildung 3 besitzt die KE *Schnittstelle* der Quell-Anforderung die Ausprägung *Zündschloss*. Diese Ausprägungsmenge unterscheidet sich von der Ausprägungsmenge *Zündschloss, Motorsteuergerät* der Ziel-Anforderung. Mit Hilfe von Ähnlichkeitsmaßen lässt sich dieser Unterschied in Prozent ausdrücken.

4.2 Ähnlichkeitsfunktionen

An dieser Stelle werden die Ähnlichkeitsfunktionen in DOORS beispielhaft angedeutet. Dies soll zugleich veranschaulichen, wie der Ansatz in der Praxis eingesetzt werden kann. Abbildung 4 zeigt einen WvT-Fall f sowie zwei WvT-Probleme x und y. Mit Hilfe der Ähnlichkeitsfunktionen soll ermittelt werden, wie ähnlich x und y zu f sind. Die ausführliche Beschreibung der Ähnlichkeitsfunktionen und ihre schrittweise Anwendung auf f und x sowie auf f und y erfolgt in [No15].

Die dargestellten Artefakte besitzen klassifizierenden Eigenschaften (KE). Die Anforderungen haben die Anforderungs-KE *Baureihe* und *Schnittstelle*. Da WvT-Probleme/Fälle eine Quell- und eine Ziel-Anforderung haben, existieren je zwei Ausprägungsmengen $Baureihe_{Quelle}$ und $Baureihe_{Ziel}$ bzw. $Schnittstelle_{Quelle}$ und $Schnittstelle_{Ziel}$. Der Testfall hat die Testfall-KE *Testziel*. Testfall-KE haben lediglich eine Ausprägungsmenge, weil nur ein Testfall an WvT-Problemen/Fällen beteiligt ist. In Anforderungswerkzeugen - wie beispielsweise DOORS - werden diese KE durch Attribute (d.h. Spalten) realisiert.

Für Ähnlichkeiten in Prozent muss die Anzahl der KE-Ausprägungen bekannt sein. Dies sind $\#KE_{Baureihe} = 3$ (M-, S-, E-Klasse), $\#KE_{Schnittstelle} = 4$ (Wish Wash Control, Türsteuergerät, Trunk Lock, Instrument Cluster) sowie $\#KE_{Testziel} = 1$ (Korrekte Schnittstellen).

Quell-Systemlastenheft (SLH_{Quell})

Quell-Anforderungen	Wv	T?	Baureihe	Schnittstelle
Q-X (WvT-Problem x)	Z-X	Ja	S	WWC TSG
Q-Y (WvT-Problem y)	Z-Y	Ja	S	WWC IC

Systemtestspezifikation (STS)

Testfälle	T-Links	Testziel
T-X (WvT-Problem x)	Q-X, Z-X	Schnittstelle
T-Y (WvT-Problem y)	Q-Y, Z-Y	Schnittstelle

Ziel-Systemlastenheft (SLH_{Ziel})

Ziel-Anforderungen	Wv	T?	Baureihe	Schnittstelle	WvT-Prüfhinweis
Z-X (WvT-Problem x)	Q-X	Prüfen	E	WWC TSG TLK	Schnittstelle geändert, Schnittstellentestfälle prüfen: T-X
Z-Y (WvT-Problem y)	Q-Y	Prüfen	E	WWC	Schnittstelle geändert, Schnittstellentestfälle prüfen: T-Y

Falldatenbank

WvT-Fälle	WvT-Entscheidung	Q: Baureihe	Z: Baureihe	Q: Schnittstelle	Z: Schnittstelle	Test: Testziel
Schnittstelle geändert, Schnittstellentestfälle prüfen (WvT-Fall f)	Prüfen	M	S	WWC TSG	WWC TSG TLK	Schnittstelle

Abb. 4: Beispielfall und -probleme

Zur Bewertung der Ähnlichkeit werden binäre Ähnlichkeitsmaße eingesetzt [Br13]. Diese basieren auf der Enthaltenseinsanalyse von zwei Mengen: Ein Element kann in beiden Mengen (a:11), exklusiv in nur einer der beiden Mengen (b:10, c:01) oder in keiner der beiden Mengen vorkommen (d:00). Zunächst wird für den in der Datenbank abgelegten Fall f die Ähnlichkeit der damaligen Quell- und Ziel-Anforderungen berechnet. Hier führt z.B. die Enthaltenseinsanalyse der Schnittstelle-KE zu dem Viertupel (a:2,b:0,c:1,d:1)

Das Tupel (a,b,c,d) fließt in ein binäres Ähnlichkeitsmaß, beispielsweise Rogers Tanimoto, ein. Dieses ist definiert als $\frac{a+d}{a+2b+2c+d}$. Mit dem Ergebnis der Enthaltenseinsanalyse der Schnittstellen-KE innerhalb von f führt wegen $\frac{2+1}{2+2*0+2*1+1}$ zu 60%. Damit sind die Schnittstellen der Quell-Anforderung von f zu 60% ähnlich zu den Schnittstellen der Ziel-Anforderung von f. Diese einzelne Ähnlichkeit wird als lokale Ähnlichkeit bezeichnet. Die globale Ähnlichkeit setzt sich aus allen lokalen Ähnlichkeiten zusammen.

Zuvor wurde die Ähnlichkeit der Anforderungen innerhalb des Falls betrachtet. Diese Änderung der Anforderungs-KE von Quelle zu Ziel wird als Trend bezeichnet. Der Trend der Schnittstellen-KE von Problem x ist wegen (a:2,b:0,c:1,d:1) 60%. Auch für Problem y ergibt sich wegen (a:1,b:1,c:0,d:2) ein Trend von 60%. Im Beispiel gilt, dass sich in f, x und y die Schnittstellen von Quelle zu Ziel gleich stark geändert haben. Die gleich starke Änderung wird mit Hilfe des Abstands $100\% - |\text{Fall-KE Trend} - \text{Problem-KE Trend}|$ bzw. $100\% - |60\% - 60\%|$ dargestellt. Die Ähnlichkeit des Schnittstellen-Trends von f und x sowie f und y liegt somit jeweils bei 100%. Auch der Trend der Baureihen-KE hat sich bei f, x und y mit jeweils (a:0,b:1,c:1,d:1) gleich stark geändert. Die Ähnlichkeit des Baureihen-Trends von f und x sowie f und y liegt damit auch bei jeweils 100%.

Weil Testfälle in WvT-Problemen/Fällen nicht als Paar auftreten, können Testfall-KE nicht trendbasiert, sondern lediglich direkt, d.h. von Falltestfall zu Problemtestfall, verglichen werden. Im Beispiel existiert lediglich ein Testziel. Damit führt der Vergleich der Testziele von f und x sowie von f und y jeweils zu (a:1,b:0,c:0,d:0) oder 100%.

Die globale Ähnlichkeit von f und x sowie f und y ergibt sich aus dem Produkt der lokalen Ähnlichkeiten. Da alle lokalen Ähnlichkeiten, d.h. die beiden Anforderungstrends und der einzelne direkte Vergleich der Testfall-KE, bei 100% liegen, liegt auch die globale Ähnlichkeit bei 100%. Der Prüfhinweis von f (Schnittstellen geändert. Schnittstellentestfälle prüfen) kann demnach sowohl für x als auch für y übernommen werden.

Im Beispiel sind die Quell- und Ziel-Schnittstellen von f und x jeweils identisch, während die Quell- und Ziel-Schnittstellen von f und y nicht identisch sind. Durch das obige Vorgehen wird dieser Unterschied nicht berücksichtigt. Neben dem trendbasierten Vergleich können auch Anforderungs-KE direkt verglichen werden. Der direkte Vergleich der Quell-Schnittstellen von f und x führt wegen (a:2,b:0,c:0,d:2) zu 100%. Im Gegensatz dazu führt der direkte Vergleich der Quell-Schnittstellen von f und y wegen (a:1,b:1,c:1,d:1) lediglich zu 33%. Neben den Quell-Schnittstellen werden auch die Ziel-Schnittstellen von f und x (100%) bzw. f und y (50%) direkt miteinander verglichen. Damit die Anforderungs-KE nicht zweimal in die globale Ähnlichkeit einfließen, wird der Mittelwert der beiden Vergleiche gebildet. Im Beispiel liegt die globale Ähnlichkeit beim direkten Vergleich aller KE von f und x bei 20%. Der Vergleich von f und y führt zu lediglich $8,4\%$ Ähnlichkeit.

Da es wahrscheinlich ist, dass sich die Baureihe zwar ändert aber der Fall andere Baureihen als das Problem enthält, können die Baureihen-KE trendbasiert und die Schnittstellen-KE sowie die Testziel-KE direkt verglichen werden. Dies führt im Beispiel für f und x zu einer globalen Ähnlichkeit von 100% während f und y zu 42% ähnlich sind. In der Falldefinition kann je nach Anwendungszweck pro Anforderungs-KE festgelegt werden, ob sie trendbasiert oder direkt verglichen werden soll (nicht in Abb. 4 dargestellt).

5 Case-Based Reasoning

Die beschriebenen Techniken zur Ermittlung von Ähnlichkeiten mit dem Ziel der Übernahme von Prüfhinweisen orientieren sich an Techniken von CBR (Case-Based Reasoning). CBR basiert auf zwei Prinzipien: (1) Ähnliche Probleme haben ähnliche Lösungen und (2) ähnliche Probleme treten immer wieder auf. Diese beiden Grundsätze lassen sich sehr einfach auf das WvT-Problem anwenden: (1) Ähnliche WvT-Probleme haben ähnliche Lösungen und (2) ähnliche WvT-Probleme treten immer wieder auf. CBR ist in vielen Domänen weit verbreitet. Häufige Anwender sind neben Softwaretechnikern beispielsweise auch Mediziner: (1) Ähnliche Krankheiten haben ähnliche Behandlungsmethoden und (2) ähnliche Krankheiten treten immer wieder auf.

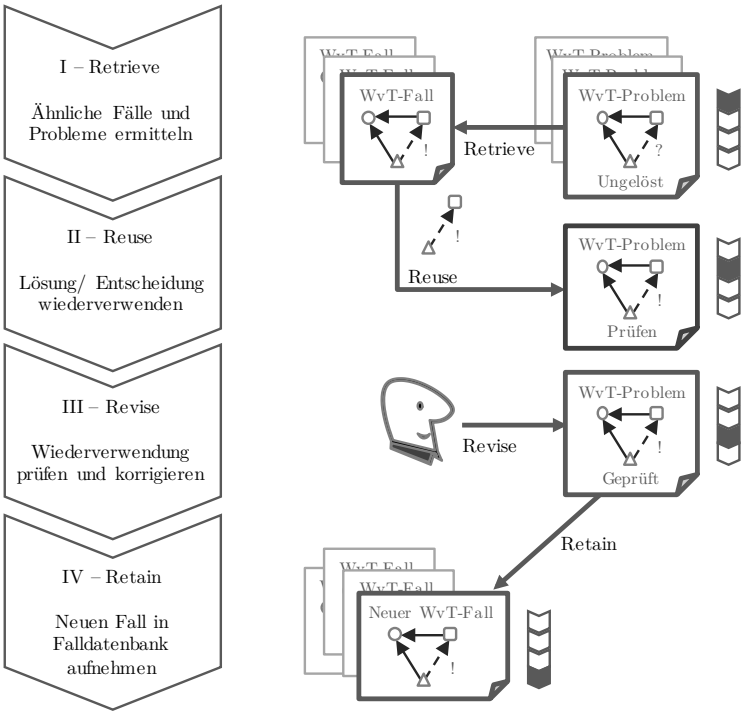


Abb. 5: CBR-Methode gemäß [AP94, S.8]

Abbildung 5 zeigt auf der linken Seite das allgemeine CBR-Vorgehen. Im ersten Schritt werden die ähnlichen Fälle für ein aktuelles Problem ermittelt (*retrieve*). Wenn ein ähnlicher Fall gefunden wurde, wird der gespeicherte Lösungsvorschlag des Falls zur Lösung des aktuellen Problems wiederverwendet (*reuse*). Im dritten Schritt wird die Lösung des aktuellen Problems geprüft und gegebenenfalls angepasst (*revise*). Im letzten Schritt wird das aktuell gelöste Problem in die Falldatenbank aufgenommen, damit es für zukünftige CBR-Zyklen zur Verfügung steht (*retain*).

Die rechte Seite von Abbildung 5 zeigt das spezifische CBR-Vorgehen. Im ersten Schritt werden die zuvor beschriebenen Ähnlichkeitsfunktionen angewendet, um für ein aktuelles WvT-Problem alle ähnlichen WvT-Fälle aus der Falldatenbank zu ermitteln (*retrieve*). Im Anschluss daran werden die Prüfhinweise der ähnlichen WvT-Fälle in das aktuelle WvT-Problem übertragen (*reuse*). Im dritten Schritt findet eine manuelle Überprüfung des zu lösenden WvT-Problems mit Hilfe der detaillierten Prüfhinweise statt (*revise*). Dabei wird auch die Korrektheit der Prüfhinweise kontrolliert und gegebenenfalls korrigiert. Im letzten Schritt werden korrigierte oder neue Prüfhinweise der gelösten WvT-Probleme in WvT-Fälle umgewandelt und in die Falldatenbank aufgenommen (*retain*).

Die Ähnlichkeitsfunktionen wurden zu Evaluationszwecken zunächst in Haskell implementiert und anschließend in DXL (DOORS eXtension Language) nachimplementiert. Damit konnten die ersten beiden CBR-Schritte in DOORS umgesetzt werden.

Die fallbasierte Filterung wurde in zwei realen Systemen erprobt. Konkrete Beispiele für fallbasierte Filterungen sind die im obigen Beispiel beschriebene Schnittstellenänderung mit beteiligtem Schnittstellentestfall sowie:

- Wenn sich die ASIL-Einstufung einer wiederverwendeten Anforderung geändert hat, dann müssen die Sicherheitstestfälle überprüft werden.
- Wenn sich der Eigentümer einer wiederverwendeten Anforderung geändert hat, dann muss der Eigentümer der Testfälle überprüft werden.

Die ersten Ergebnisse zeigen, dass es technisch möglich ist, klassifizierte WvT-Probleme und WvT-Fälle automatisch miteinander hinsichtlich der Ähnlichkeit ihrer klassifizierenden Eigenschaften zu vergleichen und einen detaillierteren Prüfhinweis zu übertragen. Da die WvT-Fälle individuell definiert werden können, sind die Möglichkeiten zur Erkennung suspekter WvT-Probleme lediglich durch die klassifizierenden Eigenschaften beschränkt. Im betrachteten Umfeld existiert eine Vielzahl solcher Attribute. Daher existieren auch zahlreiche nützliche WvT-Fälle.

Obwohl es technisch möglich ist, obige Anfragen zu stellen, ist eine volle Automatisierung des CBR-Zyklus im betrachteten Umfeld momentan schwierig: Zum einen liegt dies an den heterogenen Vorgehensweisen bei der Klassifikation von Anforderungen und Testfällen. Zum anderen bedarf es an Quality Gates immer eines manuellen Reviews. Dieses manuelle Review entspricht dem CBR-Schritt *revise*.

6 Zusammenfassung

In diesem Papier wurde ein bestehendes Verfahren zur automatischen Verlinkung von Testfällen mit wiederverwendeten Anforderungen um eine neue Komponente erweitert, mit der suspekte Verlinkungssituationen genauer charakterisiert werden können. Hierzu werden in einer Falldatenbank bereits gelöste Verlinkungssituationen als Erfahrungswissen abgelegt und diese beim Auftreten einer ähnlichen Verlinkungssituation wiederverwendet. Grundlage des Verfahrens ist eine fallbasierte Filterung, bei der anhand von klassifizierenden

Eigenschaften der betrachteten Entwicklungsartefakte entschieden werden kann, ob ein Fall der Falldatenbank zu einem aktuell zu lösenden Problem Ähnlichkeiten aufweist. Die Technik lässt sich durch vier Aktivitäten in ein vorhandenes Prozessvorgehen methodisch integrieren. Die ersten beiden Aktivitäten wurden als Erweiterung von DOORS implementiert. Die Erprobung in zwei realen Systemen belegt die generelle technische Machbarkeit des vorgestellten Verfahrens. In einer größer angelegten Feldstudie wollen wir Trade-offs hinsichtlich der Ähnlichkeitsfunktionen zur Optimierung ihrer Parameter ermitteln.

Literaturverzeichnis

[AI11] Azmi, Azri; Ibrahim, Suhaimi: Implementing Test Management Traceability Model to Support Test Documents. International Journal of Digital Information and Wireless Communications (IJDIWC), 1(1):109–125, 2011.

[AP94] Aamodt, Agnar; Plaza, Enric: Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI communications, 7(1):39–59, 1994.

[Br13] Brosius, Felix: Distanz- und Ähnlichkeitsmaße. In: SPSS 21, Kapitel 31, S. 693–709. mitp Professional, 1. Auflage, 2013.

[CHGZ12] Cleland-Huang, Jane; Gotel, Orlena; Zisman, Andrea: Software and Systems Traceability. Springer, 2012.

[Ki08] Kirova, Vassilka; Kirby, Neil; Kothari, Darshak; Childress, Glenda: Effective Requirements Traceability : Models , Tools , and Practices. BELL LABS Technical Journal, 12(4):143–157, 2008.

[NHK14] Noack, Thomas; Helke, Steffen; Karbe, Thomas: Reuse-Based Test Traceability: Automatic Linking of Test Cases and Reusing Requirements. International Journal On Advances in Software, 7(3/4), 2014.

[No13] Noack, Thomas: Automatische Verlinkung von Testfällen und Anforderungen: Testplangesteuerte Filterung. Softwaretechnik-Trends, 33(4):5–6, 2013.

[No15] Noack, Thomas: Automatische Verlinkung von Testfällen und Anforderungen - Eine wiederverwendungsorientierte Methode (eingereicht). Dissertation, TU Berlin, 2015.

[PH95] Pohl, Klaus; Haumer, Peter: HYDRA: A Hypertext Model for Structuring Informal Requirements Representations. In: Proceedings of the 2nd International Workshop on Requirements Engineering (REFSQ). S. 118–134, 1995.

[Po08] Pohl, Klaus: Requirements Engineering - Grundlagen, Prinzipien, Techniken. dpunkt.verlag, 2. Auflage, 2008.

[RWA07] Rochimah, Siti; Wan Kadir, Wan M. N.; Abdullah, Abdul H.: An Evaluation of Traceability Approaches to Support Software Evolution. In: International Conference on Software Engineering Advances (ICSEA 2007). IEEE Computer Society, S. 19–38, August 2007.

[SL05] Spillner, Andreas; Linz, Tilo: Basiswissen Softwaretest. dpunkt.verlag, 2005.

[To12] Torkar, Richard; Gorschek, Tony; Feldt, Robert; Svahnberg, Mikael; Raja, Uzair Akbar; Kamran, Kashif: Requirements Traceability: A Systematic Review and Industry Case Study. International Journal of Software Engineering and Knowledge Engineering, 22(3):385–433, 2012.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Eine Stochastische Testmethodik für den Robustheitstest Automobiler Steuergeräte

Hans-Werner Wiesbrock[1]

Abstract: Heute laufen vielzählige Applikationen auf einem Automobilen Steuergerät, greifen auf gemeinsame Speicher und Sensorik zurück und steuern unterschiedlichste Aktoren an. Sie in ihren Wechselwirkungen und zeitlichen Abhängigkeiten untereinander zu testen ist aufwendig, kostenintensiv und schwierig. Ein neuer Ansatz soll hier helfen.

Im Stochastischen Robustheitstest werden Testdaten zur Eingabe in den Prüfling über parallele Markov Automaten generiert. Der Testlauf selber wird auf die Einhaltung allgemeiner Eigenschaften wie: Blinker müssen stets vorn und hinten synchron leuchten, überprüft, eine Verletzung als Fehler protokolliert. Eine spezielle Abfragesprache, die auch erlaubte zeitliche Verzögerungen in den Erwartungen erfassen kann, unterstützt den Tester bei seiner Formulierung der Regeln. Aus dem Ansatz heraus leitet sich in natürlicher Weise auch eine Klasse von Testende Kriterien ab.

Explorativ entwickelt und gemeinsam erprobt wurde dieser Ansatz parallel zur Entwicklung eines Zentralen Bordnetz Steuergerätes bei einem großen Automobil Hersteller.

Keywords: Stochastische Testdaten Generierung, Markov Automaten, Regelbasierte Auswertung, Test der Robustheit

[1] ITPower Solutions GmbH, Kolonnenstraße 26, 10829 Berlin, hans-werner.wiesbrock@itpower.de

1 Einleitung

Zahlreiche Anwendungen laufen heute auf einem Automobilen Steuergerät und viele von ihnen interagieren asynchron über Bus Kommunikationen mit anderen Steuergeräten. So sind bspw. auf einem Zentralen Bordnetz Steuergerät Wischer, Außenlicht, Zentralverriegelung und viele andere Funktionalitäten implementiert und über mehrere CAN und LIN Busse mit anderen ECUs verbunden. Ungünstige Zeitabläufe und Nutzung gemeinsamer Ressourcen können dazu führen, dass bei gleichzeitiger Betätigung des Heckrollos und Startens des Motors das Abblendlicht kurz blinkt, der Scheibenwischer zuckt oder andere unerwünschte Seiteneffekte auftreten. Wie können wir im Test absichern, dass die vielfältigen Applikationen sich nicht gegenseitig beeinträchtigen, dass es nicht zu sporadischen Ausfällen und Fehlfunktionen in ihrer Interaktion kommt? Dass zeitliche Jitter, die zu einer merklichen Verzögerung oder gar zum Ausfall führen, nicht erst vom Kunden, sondern im Vorfeld entdeckt werden können?

1.1 Test einzelner Applikationen

Die Hersteller/Zulieferer von Steuergeräten und ihren Anwendungen haben zumeist die Hardware für sich und auch die einzelnen Funktionen umfangreich getestet. Wie die Integration sich im Verbund mit anderen Steuergeräten oder auch das komplexe Zusammenspiel vieler gleichzeitiger Anwendungen verhält kann z.T. von Ihnen nicht oder nur unzureichend getestet werden. Das liegt nicht zuletzt an dem hohen Aufwand, geeignete Testszenarien zu entwickeln und an der schwierigen Auswertung solcher Testläufe. Was ist ein Fehlverhalten, wenn vieles zugleich angesteuert wird?

1.2 Qualitätssicherung nicht-funktionaler Eigenschaften

Auch die verschiedenen Sicherheitsnormen und Qualitätsstandards fordern von heutigen Steuergeräten im Automobil, der Eisenbahn oder anderen kritischen Bereichen u.a. ihre Robustheit. In der Standard Literatur wird Robustheit bestimmt als Eigenschaft, auch in ungewöhnlichen Situationen definiert zu arbeiten und sinnvolle Reaktionen durchzuführen, [Li02]. Meist versucht man, konstruktiv dieser Forderung zu begegnen, d.h., durch Verwendung erprobter Architekturen und Designprinzipien ein robustes System zu entwickeln. Eine Teilfrage ist hier: Funktioniert mein System definiert auch bei allen möglichen Eingaben mit zeitlichen Ungewissheiten? Model Checking und andere Verifikationstechnologien versuchen, durch erschöpfende Analyse des Testeingaberaumes diese Frage zu beantworten. Doch ist dieser hoch dimensionale Raum bei realen Steuergeräten und ihren Implementierungen nicht mehr zu explorieren und die Frage nach möglicher Testunterstützung taucht auf. Auf diese Fragen versucht eine neue Testmethodik, der Stochastische Robustheitstest, eine Antwort zu geben.

1.3 Lösungsansatz

Aus den obigen Anmerkungen leiten sich zwei vordringliche Probleme ab:

- Wie erhält man mit möglichst geringem Aufwand die nötigen Testeingaben?
- Wie beurteilt man sinnvoll den Testausgang?

Auf diese beiden Fragen versucht der Stochastische Robustheitstest (SRT) Antworten zu geben. Ausgehend von Benutzer Profilen werden parallele Markov Automaten definiert, die die Testeingaben für den Prüfling erzeugen, s. auch [AL15],[se15]. Da die einzelnen Funktionen im Allgemeinen bereits gut getestet werden, siehe 1.1, ist das mögliche Fehlverhalten weniger in komplexen Funktionsabläufen zu suchen, als in einfachen Konflikten oder Verletzungen allgemein gewünschter Erwartungen. Dazu wird eine speziell für diesen Anwendungsfall entwickelte Abfragesprache vorgestellt die auch erlaubte zeitliche Verzögerungen in den Erwartungen erfassen kann. Dies ist nötig, da asynchrone Kommunikation definierte zeitliche Latenzen gestattet. Grundlegende Ideen hierzu findet man bereits in [Im07] . Im folgenden Kapitel 2 wird zunächst gezeigt, wie man beliebig viele sinnvolle Testeingaben einfach erhält. Dazu kann auf Standard Techniken (Markov Automaten) zurück gegriffen werden, die kurz eingeführt und anschließend leicht erweitert werden. Im Kapitel 3 wird dann eine Abfragesprache zur Beschreibung der Allgemeinen Erwartungen an die Testläufe und deren Beurteilung definiert, ihre theoretischen Grundlagen skizziert, sowie ihre Anwendung in Beispielen gezeigt. Ziel dieser Arbeit ist es, eine auf stochastischen Methoden beruhende Testmethode vorzustellen, die wechselseitig-bedingte Störungen effektiv aufspüren kann. Zugleich wird eine Werkzeugunterstützung gezeigt, mit der diese Methode vollautomatisch umgesetzt wurde.

Anmerkung Die mit dem SRT möglichen Prüfungen liefern keinen Ersatz für die weiterhin notwendigen umfassenden funktionalen Tests, sondern sind komplementär als Ergänzung zu den bereits bestehenden Testfällen und Testvorgehen konzipiert.

2 Markov Automaten

Nach einer kurzen Motivation wird in diesem Kapitel die Theorie der Markov Automaten skizziert und ihre spezielle Anwendung im **SRT** beschrieben.

2.1 Problemstellung

Eine zentrale Aufgabe des Tests einzelner Applikationen ist die Erstellung geeigneter Testszenarien und Eingabewerte, für Gutfälle als auch für Negativtests. Ausgehend von den Anforderungen werden zumeist typische Anwendungsfälle geprüft, Varianten von ihnen sowie einige Extreme. Im systematischen Test werden eventuell noch Code Überdeckungskriterien gemessen und ergänzende Tests abgeleitet.

Soll hingegen die gleichzeitige Ansteuerung verschiedenster Applikationen mit ihrer Kombinatorik systematisch getestet werden, so ist diese Aufgabe ungleich schwerer. Es gibt meist keine Anforderungen dazu oder nur sehr allgemein gehaltene. Lasten- und Pflichtenhefte sind häufig nach einzelnen Funktionen oder Featuren aufgeteilt und ihr Zusammenspiel nur in 'mit geltenden Unterlagen' festgehalten. Welche Kombinationen sollte man testen? Welche parallelen Szenarien? Ein rein zufälliger Test, naiv stochastische Eingaben, führen zu überwiegend inkonsistenten Werten, die bei geringster Plausibilisierung im Prüfling durch ihn bereits verworfen werden. Um zu sinnvollen, interessanten Eingaben zu kommen, sollte man sich an Benutzerprofilen orientieren. Dazu gibt es eine etablierte Technik, die Markov Automaten.

2.2 Definition von Markov Automaten

Automobile Steuergeräte sind spezielle Eingebettete Systeme und so gilt auch für sie, dass Eingabewerte in der Regel über die Zeit sequenzierte Werte sind. So ist die stochastische Erzeugung von Testeingaben im engeren Sinn Resultat eines stochastischen Prozesses. Wir wählen hierzu eine spezielle Form : Die Markov-Kette. Ohne weiter auf ihre genaue Definition [WS13, Oc05] einzugehen soll hier eine einfache Konstruktion beschrieben werden, nämlich über Markov Automaten. Dies sind endliche Automaten, deren Zustandsübergänge über Wahrscheinlichkeiten gesteuert werden.

Beispiel: Stochastische Generierung von Zündschlüsselstellungen

In diesem Beispiel werden vier verschiedene Zündschlüsselstellungen betrachtet und definiert, mit welcher Wahrscheinlichkeit die Übergänge stattfinden. Angenommen, der

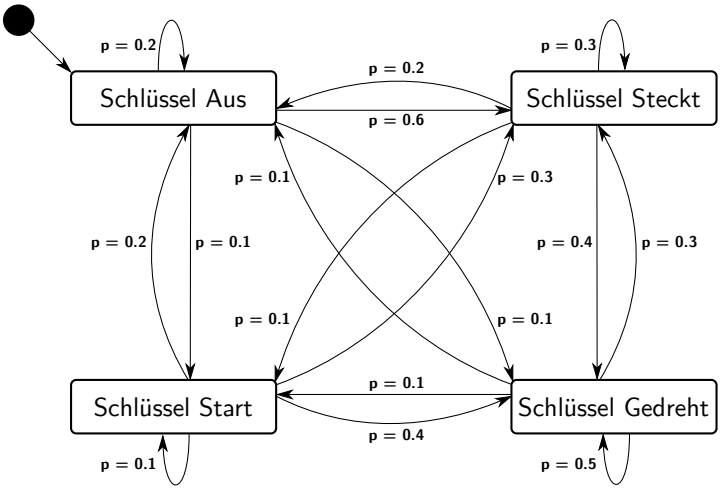


Abb. 1: Testdatenautomat: Zündung

Schlüssel befinde sich in der 'Aus'-Stellung. Dann soll bei jedem 10. mal der Schlüssel anschließend in die 'Gedreht' Stellung übergehen: *Schlüssel Aus* $\overset{[p=0.1]}{\rightarrow}$ *Schlüssel Gedreht*.

Einfacher zu beschreiben ist ein solcher Automat durch eine sogenannte Übergangs- oder Transitions-Matrix.

State	Aus	Steckt	Gedreht	Start
Aus	0,2	0,2	0,1	0,2
Steckt	0,6	0,3	0,3	0,3
Gedreht	0,1	0,4	0,5	0,4
Start	0,1	0,1	0,1	0,1

Tab. 1: Übergangswahrscheinlichkeiten

Gegeben solch ein Markov Automat 'würfelt' man jetzt bspw. alle 100 [ms] (Periode) und wechselt entsprechend der 'Würfelzahl' und der angegeben Wahrscheinlichkeit der Transition von einem Zustand in den nächsten. Als Ergebnis erhält man einen stochastischen Prozess, der nach und nach verschiedene Zustände des Automaten durchläuft. Technisch realisiert wird ein solcher Automat durch Pseudozufallsvariablen unter Verwendung eines Seed Wertes, um auf diese Weise reproduzierbare Läufe zu erhalten.

2.3 Test Semantik der Wahrscheinlichkeiten

Ohne zu sehr auf die mathematischen Nachweise hier einzugehen, soll kurz erläutert werden, wie die Übergangswahrscheinlichkeiten aus Testersicht zu lesen sind und wie geeignete Werte ermittelt werden können. Durch eine Rückführung auf ein Bernoulli Experiment lässt sich leicht ermitteln, wie groß die durchschnittliche Verweildauer in einem Zustand ist, in diesem Beispiel:

State	Aus	Steckt	Gedreht	Start
Mittlere Verweildauer	1,25	1,43	2	1,1

Tab. 2: Mittlere Verweildauer in [Periode] Einheiten

Die Verhältnisse der Übergangswahrscheinlichkeiten zu anderen Zuständen hingegen steuern, wie häufig ein spezieller Folgezustand gewählt wird:

Schlüssel Aus $\overset{[p=0.2]}{\rightarrow}$ *Schlüssel Steckt* und *Schlüssel Aus* $\overset{[p=0.1]}{\rightarrow}$ *Schlüssel Gedreht*

besagt, dass doppelt so häufig von 'Aus' nach 'Steckt' gegangen wird als nach 'Gedreht'.

Umgekehrt gilt aber auch folgendes. Gegeben seien alle relativen Übergangszahlen zu anderen Zuständen, im Beispiel:

Schlüssel Aus$\rightarrow$ *Schlüssel Steckt* doppelt so häufig wie *Schlüssel Aus*$\rightarrow$ *Schlüssel Gedreht*

und ebenso für die anderen Zustandsübergänge von *Schlüssel Gedreht*$\rightarrow$ *Schlüssel Start* et cetera. Außerdem seien die gewünschten mittleren Verweildauern in Einheiten der 'Würfel' Periode bekannt. Dann ist dadurch eindeutig die gesamte Übergangsmatrix bestimmt.

Wir können also aus Benutzerprofilen für die einzelnen Applikationen leicht die entsprechenden Transistionsmatrizen ableiten!

Die Matrixelemente (Übergangswahrscheinlichkeiten) sind alle positiv, der Eigenvektor zum maximalen Eigenwert ist also eindeutig (Perron Frobenius) und lässt sich einfach berechnen. Normiert ergibt dieser die mittleren Aktivierungen der einzelnen Zustände, also wie häufig in Relation zu den anderen, ein Zustand im durchschnittlichen Testlauf aktiv war.

2.4 Stochastische Generierung bei kontinuierlichen Testdaten

Das in 2.2 beschriebene Vorgehen funktioniert sehr gut, wenn die möglichen Eingabedaten aufzählwertig sind wie z.B. Schlüssel, Wischerhebel und Lichdrehschalter Stellungen. Im Falle kontinuierlicher Eingabedaten wie der Geschwindigkeit muss dieser Ansatz erweitert werden. Dazu kann man nach der Klassifikationsbaummethode vorgehen, [GG93]. Nach dieser werden die möglichen Eingabewerte in Äquivalenzklassen eingeteilt. Die zugrundeliegende Relation wird bestimmt durch die Uniformitätshypothese, d.h., es werden Eingabedaten als annähernd gleich klassifiziert, wenn sich die Applikation bei ihrer Eingabe voraussichtlich annähernd gleich verhält. Der kontinuierliche Wertebereich wird auf diese Weise in eine endliche Zahl von Klassen ($\rightarrow$ Zuständen im Markov Automat) partitioniert. Damit ist man in einer ähnlichen Situation wie im Falle enumerationswertiger Eingaben. Doch sind hier die Transitionen etwas allgemeiner zu definieren.

Beispiel: Stochastische Generierung von Geschwindigkeitseingaben

Sei die Geschwindigkeit eines Fahrzeuges ein Eingabedatum einer Applikation. Gemäß der Klassifikationsbaum Methode identifiziert man verschiedene Geschwindigkeitsbereiche [0-5] km/h, [5-50] km/h und [50-150] km/h. Nach Kapitel 2.3 wird eine Übergangsmatrix für die Bereiche definiert, siehe linke Seite in Abb. 2.

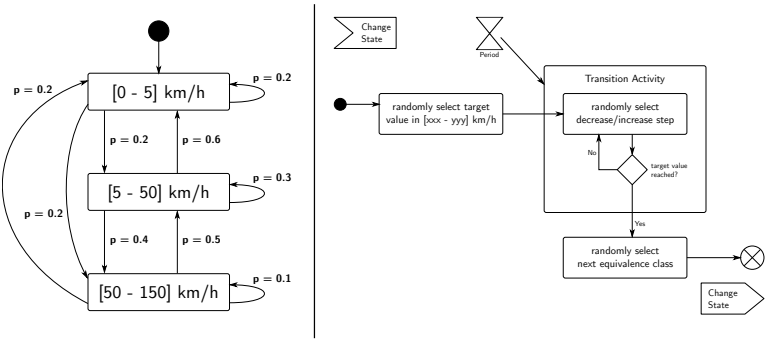


Abb. 2: Vorgehen bei kontinuierlichen Daten

In Erweiterung zum Vorgehen in Kapitel 2.2, wählt man zunächst zufällig einen Zielwert in der nächsten Äquivalenzklasse, z.B. 37 km/h $\in$ [5,50] km/h. Gegeben sei nun für jede Äquivalenzklassen Transition eine Schrittweite und eine Wahrscheinlichkeit, z.B. 0.5

km/h, und p = 0.1. Dann würfelt man mit Wahrscheinlichkeit p erneut, ob zur aktuellen Größe die Schrittweite hinzu addiert, oder subtrahiert wird. Man erhält so einen Random Walk mit definierter Schrittweite, der abhängig von der Wahrscheinlichkeitsangabe schneller oder langsamer den Zielwert erreicht. Ist er erreicht wird erneut ein Zielwert ausgewürfelt und das Verfahren beginnt von vorne.

Allgemein kann man so zu kontinuierlichen Eingabegrößen mit Hilfe der Klassifikationsbaummethode geeignete Markov Automaten assoziieren und entsprechende Übergänge modellieren.

Auf diese Weise lassen sich dezidiert die Testdaten für beliebig viele Applikationen über parallel laufende Markov Automaten stochastisch definieren und generieren.

2.5 Überdeckungskriterium

Bei dieser Testmethode werden die Testdaten wie in 2.2 und 2.4 beschrieben über spezielle Zustandsautomaten (Markov Automaten) generiert. Ziel des **SRT** ist es, gegenseitige Störeinflüsse auf dem Steuergerät zu entdecken. Insbesondere sollten deshalb in den Tests vielfältige Ansteuerungskombinationen durchlaufen werden. Als Kriterium für die Testtiefe bietet sich an, die Anzahl der unterschiedlichen Kombinationen in den Ansteuerungen zu verwenden. Dazu werden folgende Überdeckungsmaße vom Grad $k \in \mathbb{N}$ vorgeschlagen:

$$CovM(k) \stackrel{def}{=} \frac{k!(n-k)!}{n!} \sum_{n-tuple(S_1,S_2,..,S_k)} \frac{\sharp\text{activated Substates in } (S_1 \times \cdots \times S_k)}{\sharp\text{Substates in } (S_1) \cdots \sharp\text{Substates in } (S_k)}$$

mit S_i - Markov Automat, $\sharp(S_i)$ - Anzahl der Zustände in S_i, $S_1 \times S_2$ - Produkt Automat, n-Anzahl aller Markov Automaten.

Beispiel: CovM(2)

Zu jedem Paar von Testdatengeneratoren wird festgehalten, welche ihrer Zustandskombinationen in den Testläufen angenommen wurden. Die Anzahl aller Möglichkeiten für ein Paar ist offensichtlich gleich dem Produkt aus der Anzahl ihrer Unterzustände. Es wird nun die Summe aller durchlaufenen Zustandskombinationen, gemittelt durch die Anzahl aller möglichen, gebildet. $\frac{n!}{2!(n-2)!}$ *ist die Anzahl aller Paare. Das Überdeckungsmaß CovM(2) wird also durch den Divisor auf 1 normiert. Ist CovM(2) = 1, so ist zu zwei beliebigen Eingabeautomaten jede der wechselseitigen Eingabekombinationen mindestens einmal durchlaufen worden.*

Abhängig von der gewünschten Testtiefe im **SRT** lassen sich dann über die Wahl des k in CovM(k) geeignete zu erreichende Kennzahlen ableiten.

3 Überprüfung allgemeiner Eigenschaften

Wenden wir uns der zweiten Frage zu: Wie beurteilt man sinnvoll den Testausgang?

Bei allgemeinen, stochastischen Eingaben ist nicht offensichtlich, wie der Testausgang beurteilt werden soll. Die Erfahrung zeigt jedoch, dass viele der im Feldversuch aufgetretenen Fehler nicht komplexe Reaktionen auf spezifische Eingaben waren, sondern in der Regel die Verletzung einfach zu überprüfender, allgemeiner Eigenschaften wie z.B. mangelnde Synchronisation der Blinkfunktionen, fehlerhafte, sporadische Ansteuerung verschiedener Leuchten. Der Aufwand, solche Fehler frühzeitig zu entdecken, liegt demnach nicht so sehr in der Spezifikation eines ausgefeilten Testorakels, sondern in der geeigneten komplexen Ansteuerung des Systems mit der parallelen Überwachung auf die Einhaltung allgemeiner Eigenschaften! Es sind Regeln wie:

- Die beiden diskreten Ausgänge zu den Blinkleuchten müssen synchron angesteuert werden! (erlaubte Verzögerung 160 [ms])
- Die Rückfahrlichter dürfen nur bei Zündung an und Rückwärtsgang leuchten! (erlaubte Verzögerung 400 [ms])
- Nebelleuchten dürfen nicht an sein, wenn Lichtdrehschalter nicht auf Nebellicht oder Nullstellung oder AFL steht! (erlaubte Verzögerung 100 [ms])

Zu berücksichtigen ist hierbei, dass bedingt durch zwischengeschaltete Messkaskaden, CAN Bus Kommunikation usw. systematische Zeit- und Wertejitter bei den Messwerten und Ansteuerungen zu erwarten sind. Die Regeln müssen also 'bis auf definierte Zeitverzögerungen und Wertedifferenzen', sogenannte Latenzen, eingehalten werden.

3.1 Berücksichtigung zeitlicher und amplitudenmäßiger Latenzen

Betrachten wir zunächst die Forderung nach synchronem vorderen und hinterem Blinken und ihre Überprüfung im Hardware-in-the-Loop Testschrank oder im Fahrzeug. Um einen sinnvollen Vergleich von Signalen durchzuführen, sind gewisse Toleranzen in Zeit und Wert zu erlauben. (Unterhalb von 30 [ms] z.B. ist eine Abweichung der Synchronizität durch den Menschen nicht wahrnehmbar, siehe [P6] .) Kontrollierte Abweichungen in den Wertemessungen zu bewerten ist Standard und kann über verschiedene Wertvergleichskriterien abgedeckt werden. Problematisch ist jedoch die mögliche Verschiebung in der Zeit zu beurteilen. In [GW10] wird ein Konzept einer temporalen Logik mit Latenzzeiten eingeführt. In ihr sind innerhalb bestimmter Latenzzeiten keine eindeutigen Aussagen über wahr und falsch zu treffen. In dieser Arbeit wird auch gezeigt, wie dieser Ansatz zu bekannten Temporal Logiken in Relation steht, [MP92].

Ohne weiter auf diese Konzepte einzugehen, soll hier nur ein grundlegendes Konstrukt vorgestellt werden, die Implikation mit zeitlichen Latenzen:

Seien $\mathbb{A}(t)$ und $\mathbb{B}(t)$ zwei Flag Signale, d.h. boolwertige Funktionen über der Zeit t. Seien ferner Δ_F und Δ_P Latenzzeiten [ms]. Dann definieren wir für jeden Zeitpunkt t:

$$\begin{aligned} IMPLIES(\mathbb{A},\mathbb{B},\Delta_F,\Delta_P)\ (t) \quad &:\Leftrightarrow \\ &\mathbb{A}(t) = true \Rightarrow \exists t' \in [t-\Delta_F, t+\Delta_P] : \mathbb{B}(t') = true \wedge \\ &\mathbb{B}(t) = false \Rightarrow \exists t'' \in [t-\Delta_F, t+\Delta_P] : \mathbb{A}(t'') = false. \end{aligned}$$

Synchronizität mit Latenzen wird dann als zweifache Implikation definiert:

$$\begin{aligned} SYNCH(\mathbb{A},\mathbb{B},\Delta_F,\Delta_P)\ (t) \quad &:\Leftrightarrow \\ &IMPLIES(\mathbb{A},\mathbb{B},\Delta_F,\Delta_P)(t) \wedge IMPLIES(\mathbb{B},\mathbb{A},\Delta_F,\Delta_P)(t) \end{aligned}$$

Wählt man insbesondere $\Delta_P = 0$ resp. $\Delta_F = 0$ so lassen sich hierüber auch kausale Beziehungen wie: $\mathbb{A}$ verursacht $\mathbb{B}$, bzw. $\mathbb{B}$ muss $\mathbb{A}$ vorausgegangen sein, formulieren. Auch kann eine abgeschwächte Form der Transitivität für die Implikation und Synchronizität gezeigt werden, bei der die Latenzbereiche zu summieren sind.

Beispiel: Synchrones Blinken

Der vordere Blinker soll synchron zum hinteren Blinker leuchten. Kleine zeitliche Verschiebungen unterhalb von 50 ms können toleriert werden, da es die gleichzeitige Wahrnehmung nicht beeinträchtigt.

Bezeichne $U_V(t)$ die Spannungsansteuerung des vorderen Blinkers über die Testzeit, analog $U_H(t)$ für den hinteren Blinker. Die Latenzen sind dann $U(t) := [t-50ms, t+50ms]$ und die Regel lautet, dass zu allen Zeitpunkten

$$SYNCH((U_V(t) > 7{,}5V), (U_H(t) > 7{,}5V), 50, 50)(t) = true$$

gilt. Wir können diese Eigenschaft für jeden Testlauf jetzt algorithmisch überprüfen.

In [GW10] werden weitere Konstrukte aus den obigen Konzepten abgeleitet, über die Eigenschaften wie: Frühestens 30 s nach Ziehen des Schlüssels darf der CAN Bus zur Ruhe gehen, formuliert und überprüft werden können.

Konstrukt	**Beschreibung**
FOR_AT_LEAST($\mathbb{A};\Delta_t$)	Wenn $\mathbb{A}$ aktiv wird (=true), dann mindestens für eine Dauer Δ_t
	Beispiel: Mindestblinkdauer
HOLDS_AT_MOST($\mathbb{A};\Delta_t$)	Wenn $\mathbb{A}$ aktiv wird (=true), dann höchstens für eine Dauer Δ_t
	Beispiel: Maximale Blinkdauer
CHANGE_BEFORE($\mathbb{A}$**,incr;**Δ_t)	Nach einem Flagwechsel ($\mathbb{A} : false \rightarrow true$) für Δ_t =true
	Beispiel: Spätestens nach Zündungswechsel...
&& , **||**, **!**	Elementar Boolesche Verknüpfungen von Flagsignalen
	Beispiel: Zündung an und Wischerhebel betätigt...

Tab. 3: Weitere Sprachkonstrukte zur Formulierung von Regeln

3.1.1 Allgemeine Eigenschaften

Um nun Kriterien für die Beurteilung der Testausgänge zu formulieren, wird ein Messergebnis über einfache Vergleichsoperationen ($<, \leq, >, \geq, ==$) zu wahr oder falsch abstrahiert. Man kann sich so bei der Formulierung allgemeiner Regeln ohne Einbußen auf boolwertige Signale beschränken.

Man zeigt leicht, dass die in Tab. 3 angeführten Konstrukte und auch IMPLIES, resp. SYNCH die Kategorie der boolwertigen Signale über der Zeit nicht verlassen. Insbesondere kann man die verschiedenen Konstrukte ineinander verschachteln, komponieren! Auf diese Weise lassen sich beliebig komplizierte neue Boolesche Signale über der Zeit definieren.

Anmerkung: In Experimenten hat man eine definierte Startzeit des Testlaufs. Will man die Konstrukte auf Boolesche Signale über $[0, \infty[$ beschränken, ist noch das Problem der Randbedingungen zu beachten. Es soll aber an dieser Stelle nicht weiter darauf eingegangen werden.

Die Auswertung der stochastisch getriebenen Testläufe geschieht nun wie folgt. Es werden mit Hilfe der Messsignale eines Testlaufes geeignete Boolesche Signale definiert, sogenannte Regeln, die stets wahr sein sollen:

- SYNCHRON (bStBlinkerVREin, bStBlinkerHREin ; 160 [ms])
  Synchronität der beiden diskreten Ausgängen zu den Blinkleuchten. (erlaubte Verzögerung 160 [ms])
- IMPLIES((bKlemme15 && bRueckwaertsGang) , rRueckfahrlichtR ; 400 [ms])
  Die Rückfahrlichter können nur bei Zündung an und Rückwärtsgang leuchten. (erlaubte Verzögerung 400 [ms])
- IMPLIES(!(bStLdsNebellicht || bStLdsNullstellung || bStLdsAssistenzfahrlicht)) , rNebellichtR == 0 ; 400 [ms])
  Nebelleuchten nicht an, wenn Lichtdrehschalter nicht auf Nebellicht oder Nullstellung oder AFL. (erlaubte Verzögerung 400 [ms])

Diese Regeln (Boolschen Signale) werden nun parallel während des Testlaufes überwacht. Ein Fehlverhalten zur Zeit t ist dann gegeben, wenn das Signal zur Zeit t den Wert false annimmt.

Beispiel: Synchrone Ansteuerung der Blinkleuchten *Falls zu einem Zeitpunt t_0 gilt: SYNCHRON (bStBlinkerVREin, bStBlinkerHREin ; 160 [ms]) (t_0)=false, so laufen die Blinkleuchten nicht synchron!*

Aufgrund der Latenzen in der Zeit ist für eine Auswertung zu beachten, dass ggfs. der Zeitpunkt der Fehlerdetektion, im Beispiel: t_0, definiert verzögert zum eigentlichen Fehlverhalten sein kann!

In der explorativen Studie parallel zur Entwicklung eines Zentralen Bordnetz Steuergerätes konnte gezeigt werden, dass die Mächtigkeit der angebotenen Sprachkonstrukte hinreichend war, um die allgemeinen Regeln zur Zentralverriegelung, Außenlicht Ansteuerung etc. zu erfassen.

4 Realisierung

Umgesetzt wurde diese Testmethode mit Hilfe einer Werkzeug Implementierung in *C#*. Ausgehend von einer Übergangsmatrix, vorgegebenen Periode und Seed (Anfangsbedingung für die Erzeugung von Pseudo Random Variables) werden dann zyklisch stochastische Testdaten zu einem Markov Automaten generiert. Die Auswertungsalgorithmen basieren auf fest definierten zyklischen Erfassungszeiten der Messungen. Über eine geeignete GUI können die verschiedenen Aufgaben im Rahmen des **SRT** zentral bearbeitet werden.

Für jede Applikation lässt sich dann ein Markov Automat definieren. Die Zustände erlauben dann die Definition von Entry resp. During Aktionen, in denen beliebig viele Signale sequentiell gesetzt werden können, z.B. Zündung an: $Kl\_S = 1, Kl\_15 = 1, Kl\_X = 1, Kl\_50 = 0$; Die Übergangswahrscheinlichkeiten können dann direkt eingegeben oder durch relative Gewichtungen und Angabe von durchschnittlichen Verweildauern definiert werden.

Die Spezifikation von Regeln ist aufgrund der temporal-logischen Möglichkeiten nicht trivial und bedarf der Übung. Zur Vereinfachung wurden Muster Bibliotheken erstellt, so dass der Benutzer nur noch geeignete Parameter zu ersetzen hat. Auch wird ihm ein Editor angeboten, der die Syntax der Abfragesprache unterstützt. Trotzdem sollten bei der Formulierung **SRT** Experte und Anwendungsspezialist Hand in Hand arbeiten.

Um die Testdaten projektübergreifend wieder verwenden zu können, wurde eine Abstraktionsschicht implementiert, in der Signalbezeichner projektspezifisch angepasst werden können ohne die Definition der Automaten oder Regeln ändern zu müssen. Sie lassen sich dann mit logischen Signalbezeichnern formulieren. Da die Regeln und Ansteuerungen sehr allgemein sind, können so erfahrungsgemäß viele Artefakte projektübergreifend wiederverwendet werden!

Anbindung an eine Testumgebung

Die im **SRT** erstellten Testdatengenerator-Automaten setzen auch der COM Schnittstelle von CANoe auf. Über IO Karten (VT 1040, VT 2040 etc.) wird die Hardware angesteuert.

Für die Überwachung der Regeln wird die .Net Implementierung der Auswertungsalgorithmen als spezieller Simulationsknoten eingebunden. Änderungen in den Messdaten werden über CANoe Trigger an die Algorithmen weitergeleitet und direkt im Testlauf ausgewertet.

Für die Auswertung können die grafischen Möglichkeiten von CANoe genutzt werden, und die eingebaute Auswertungskomponente .

4.1 Erfahrungen und Ausblick

In der Erprobungsstudie wurde eine etwas abweichende Architektur verwendet, in der die verschiedenen Komponenten aus Matlab Simulink® generiert und als CANoe Simulationsknoten eingebunden wurden.

Prüfling war ein Zentrales Bordnetz Steuergerät. Es besitzt mehrere Hundert HW Eingangs- und Ausgangsschnittstellen, ist an verschiedenen CAN und LIN Bussen angeschlossen und zeichnet sich dadurch und einer extrem hohen Anzahl verschiedener auf ihm laufenden Funktionen gegenüber anderen Automobil Steuergeräten aus.

Grundlage für die Erstellung der Benutzerprofile waren Interviews mit den Applikationsexperten sowie die zugehörigen Lastenhefte. Es wurden ca. 25 Markov Automaten entwickelt, die im Testlauf parallel aktiviert wurden. (Funktionen: Zündung, Außenlicht, Wischeransteuerung, Zentralverriegelung,...)

Am schwierigsten gestaltete sich die Erstellung der Regeln. Hierzu wurden Funktionsanforderungsexperten und erfahrene Funktionstester in mehreren Schleifen befragt. Es wurden an die 50 Regeln aufgestellt und parallel im Testlauf überwacht.

Die Testläufe wurden über mehrere Stunden durchgeführt. Zu dem Zeitpunkt waren noch nicht Testende Kriterien implementiert, so dass keine Aussagen über die Abdeckung der Automaten gemacht werden kann.

In der Studie konnten bisher vorrangig zwei Arten von Fehlern entdeckt werden:

- Konkurrierende Zugriffe auf Treiber (Aufblitzen und mangelnde Synchronizität bei Leuchten)
- mangelnde zeitliche Abstimmung von interagierenden Applikationen

die sporadisch auftraten! Über die Anzahl der gefundenen Fehler darf ich keine Aussagen machen.

Weitere positive Effekte waren verbesserte Anforderungsdokumente. In den intensiven Interviews mit den Applikationsexperten für die korrekte Aufstellung der Regeln wurden noch einmal die grundlegenden Prinzipien und Leitlinien der verschiedenen Anwendungen diskutiert. Dies führte im Folgenden auch zu klareren Spezifikationen der Anforderungen.

Bei der Einführung und Entwicklung des **SRT** zeigte sich auch, dass eine neue Testrolle einzuführen ist: Ein **SRT** Experte, der vor allem die Abfragesprache gut beherrscht. Er hat zum einen mit den Anwendungsexperten die Regeln inhaltlich abzuklären, zum anderen mit den Testexperten die konkrete Umsetzung am Prüfstand zu besprechen.

Für die Weiterentwicklung ist vorgesehen:

- Erweiterungen der Testdatengenerierung zur Szenarien Generierung (Beispiel: Normallauf, Ruhe, Extremlauf), bei der Gruppen von Markov Automaten in sogenannten Szenarien zusammengefasst werden.

- Verbesserte Fehleranalyse Unterstützung (Getriggertes Logging, grafische Unterstützung,..)

Dringlichste Aufgabe ist jedoch eine Bewährung dieses explorierten Ansatzes als neue, ergänzende Qualitätssicherungsmaßnahme. Dazu wurde die Neuentwicklung an einen speziellen Teststand angebunden und wird ab Sommer dieses Jahres in einer laufenden Steuergeräte Entwicklung eingesetzt.

Danksagung Diese Arbeit wäre ohne die zahlreichen Diskussionen und die tatkräftige Unterstützung der Herren R. Immig, St. Laufmann und M. Wagener nicht denkbar gewesen. Ihnen sei an dieser Stelle herzlichst gedankt!

Literatur

[AL15] ALL4TEC: MaTeLo Overview. http://www.all4tec.net/MaTeLo/matelo-features.html, 15.04.2015.

[GG93] Grochtmann, Mathias; Grimm, Klaus: Classification-trees for partition testing. Software Testing, Verification and Reliability, 3(3):63–82, 1993.

[GW10] Grossmann, J.; Wiesbrock, H.-W.: An Approach and Extension of Temporal Logic for Testing Real Time Properties of Embedded Systems. unpublished manuscript, 2010.

[Im07] Immig, Ralf: Abdeckungskriterien für den funktionalen Test eines elektronischen Steuergerätes. Diplomarbeit, HU Berlin, Juni 2007.

[Li02] Liggesmeyer, Peter: Software-Qualität. Spektrum Akademischer Verlag Heidelberg-Berlin, 2002.

[MP92] Manna, Z.; Pnueli, Amir: The Temporal Logic of Reactive and Concurrent Systems. Springer, New York, 1992.

[Oc05] Ocone, Daniel: Chapter 5: Markov Chain Models. http://faculty.oxy.edu/angela/prior2005. Vorlesungsskript.

[P6] Pöppel, Ernst: Der Rahmen. Carl Hanser Verlag, München, 2006.

[se15] sepp.med: MBTsuite - The testing framework https://www.seppmed.de/produkte/mbtsuite.html, 15.04.2015.

[WS13] Waldmann, Karl-Heinz; Stocker, U. M.: Stochstische Modelle. Springer-Verlag, Kapitel Markov-Ketten, S. 9 – 57, 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

AmbiTune: Kontextorientierte Musikeinspielung im Fahrzeug

Patrick Helmholz[1], Susanne Robra-Bissantz[2]

Abstract: Die Erfassung der Umwelt durch Sensoren und die dadurch ermöglichte kontextorientierte Anpassung von Informationen und Unterhaltung wird bei mobilen Applikationen immer wichtiger. Besonders das Hörverhalten von Musik wird von der aktuellen Situation und Umwelt beeinflusst. Kontextorientierte Musikempfehlungssysteme können dabei helfen, automatisch die richtige Musik zur richtigen Zeit am richtigen Ort eingespielt zu bekommen. Speziell im Fahrzeug hat der Fahrer wenig Möglichkeiten, mit dem System zu interagieren, um die aktuell gewünschte Musik auszuwählen. Der vorliegende Beitrag zeigt anhand von Studienergebnissen und eines entwickelten Prototypen auf, wie im Rahmen des Projekts AmbiTune die kontextorientierte Musikeinspielung im Fahrzeug untersucht und ermöglicht wird.

Keywords: Fahrzeug, Kontext, Kontextorientierung, Musikempfehlung.

1 Einleitung

Individuell zugeschnittene Dienstleistungspakete rund um das Automobil gewinnen als Differenzierungsmerkmal beim Autokauf eine immer größere Bedeutung. Zukünftig werden 80 Prozent der Automobilinnovationen softwarebasiert sein [RKR07]. Moderne Fahrzeuge verfügen zunehmend über Sensorik zur Erfassung der Umwelt sowie über einen Zugang zum Internet als Quelle für weitere Daten [Ga12, Ge14]. Diese Kontextdaten werden im Fahrzeug bisher jedoch nur geringfügig genutzt, um Informations- und Unterhaltungsdienste anzupassen. Eine Bedarfsanpassung in Form von Personalisierung und Kontextbezug nimmt bei mobilen Diensten jedoch eine immer wichtigere Rolle ein. Das Marktforschungsunternehmen Gartner prognostiziert in der Bedürfnisdeckung der mobilen Nutzer in unterschiedlichen Kontexten und Umgebungen einen zukünftigen Schwerpunkt und gibt *kontextorientierte Systeme* und *Computing Everywhere* als zwei der zehn wichtigsten strategischen Technologietrends für das Jahr 2015 aus [Ga14].

Gerade im Bereich der Musikdienste wird der Kontext auf Applikationsebene bisher nur geringfügig betrachtet und eingesetzt. Die immer häufiger genutzten personalisierten Musikstreamingdienste wie Spotify oder Deezer verwenden lediglich Profilinformationen des Hörers zur kollaborativen Ermittlung ähnlicher Profile, um die

[1] TU Braunschweig, Lehrstuhl Informationsmanagement, Mühlenpfordtstr. 23, 38106 Braunschweig, p.helmholz@tu-bs.de
[2] TU Braunschweig, Lehrstuhl Informationsmanagement, Mühlenpfordtstr. 23, 38106 Braunschweig, s.robra-bissantz@tu-bs.de

Musik an den Musikgeschmack anzupassen. Oder sie ermöglichen die Angabe der Stimmung, um dementsprechend eine Playlist zu generieren. Keiner dieser Dienste verwendet bisher automatisch erfasste Kontextinformationen (z.B. Ort, Uhrzeit oder Wetter), um die Musikeinspielung anzupassen.

Im mobilen Umfeld beeinflusst der Kontext jedoch häufig den Musikwunsch des Konsumenten und besitzt aufgrund ständiger Veränderungen einen hohen Stellenwert. Ohne Einbezug des Kontextes entstehen folglich häufig Musikempfehlungen, die für den Musikhörer von geringem Nutzen sind. [KR12]

2 Bedeutung von Musik im Fahrzeug

Studien zeigen, dass das Autoradio das wichtigste Ausgabemedium für Musik ist. In einer Studie des Statistischen Bundesamtes gaben 70 Prozent der Befragten an, das Autoradio zur Musikausgabe zu nutzen. Es hat den höchsten Wert vor dem klassischen Radiogerät (ca. 67 %). Über 80 Prozent dieser Personen hören zudem täglich Autoradio. [St11a, St12] Jeden Morgen hören beispielsweise ca. 9 Millionen Autofahrer in Deutschland in ihrem Fahrzeug Radio [Ra13].

Die Musik kann während der Autofahrt verschiedene Funktionen erfüllen. Sie kann verwendet werden, um ein Gemeinschaftsgefühl bei Alleinfahrten zu erzeugen, um eine Ablenkung auf einer langen und langweiligen Fahrt zu ermöglichen, um die eigene Stimmung zu regulieren, oder einfach nur zur Unterhaltung und Ablenkung während der Fahrt [De11, DW07, Va13]. Zudem zeigen Studien, dass die Musikeinspielung während der Fahrt auch die Fahrleistung sowohl positiv als auch negativ beeinflussen kann [Br01, CM10, DW07, NH99, USE12].

Die vorgebrachten Fakten belegen, dass Radio hören eine wichtige Nebentätigkeit während der Autofahrt ist. Viele Musikkonsumenten wollen jedoch aus unterschiedlichen Gründen ihre eigene Musik hören und nicht dem klassischen Radioprogramm folgen [NHH04] und nutzen daher ihre (Online-)Musikmediatheken. In den letzten Jahren stellt dies immer mehr Musikkonsumenten vor eine große Herausforderung. Für den mobilen Bereich sind die Musikmediatheken zu umfangreich und dadurch schwer überschaubar geworden. Die Nutzer können bis zu 30 Millionen Songs von mehr als sieben Millionen Künstlern aus den verschiedensten Genres (z.B. Pop, Rock, Klassik, Techno) wählen [St14, Th11]. Hierdurch erweist sich die Auswahl eines für den aktuellen Fahrtkontext passenden Musiktitels als äußerst zeitaufwändig [KR12] und erfordert zudem Interaktion mit dem System[3] [MTL11]. 61 Prozent der deutschen Autofahrer fühlen sich durch das System inzwischen sogar abgelenkt [CS14]. So können sich die Musikkonsumenten in bestimmten Situationen nicht mehr vollständig auf ihre primäre Aktivität, nämlich der Führung des Fahrzeugs, konzentrieren. Eine

[3] Als System sind in diesem Fall sowohl das Infotainmentsystem des Fahrzeugs wie auch externe Geräte (z.B. mobile Musikplayer oder Smartphones) gemeint.

geeignete (Vor-)Auswahl der passenden Musik durch das System ist somit aus mehreren Gründen gefordert. [KR12] Hier können Musikempfehlungssysteme helfen, welche die aktuelle Situation des Fahrers und das Umfeld des Fahrzeugs berücksichtigen.

3 Kontextorientierte Musikempfehlungssysteme

Erfolgt kurzfristig keine Änderung der Präferenzen eines Nutzers, funktionieren Empfehlungssysteme, die aus dem zuvor beobachteten Verhalten auf das künftige Verhalten schließen. In vielen Bereichen und insbesondere beim Hören von Musik im mobilen Umfeld – wie bei der Autofahrt – spielt jedoch der Kontext eine entscheidende Rolle. Die Aktivität, die Umgebung oder die Stimmung eines Nutzers kann beispielsweise kurzfristig zur Veränderung der Musikpräferenzen führen. [Ba10]

Um für einen Nutzer das bestmögliche Ergebnis zu erzielen, sollten Empfehlungssysteme sowohl die Präferenzen des Nutzers und die Eigenschaft der Musikstücke, als auch den jeweiligen Kontext berücksichtigen (siehe Abbildung 1). Somit werden nicht nur personalisierte, sondern auch der aktuellen Situation entsprechende Empfehlungen gegeben. [SH08]

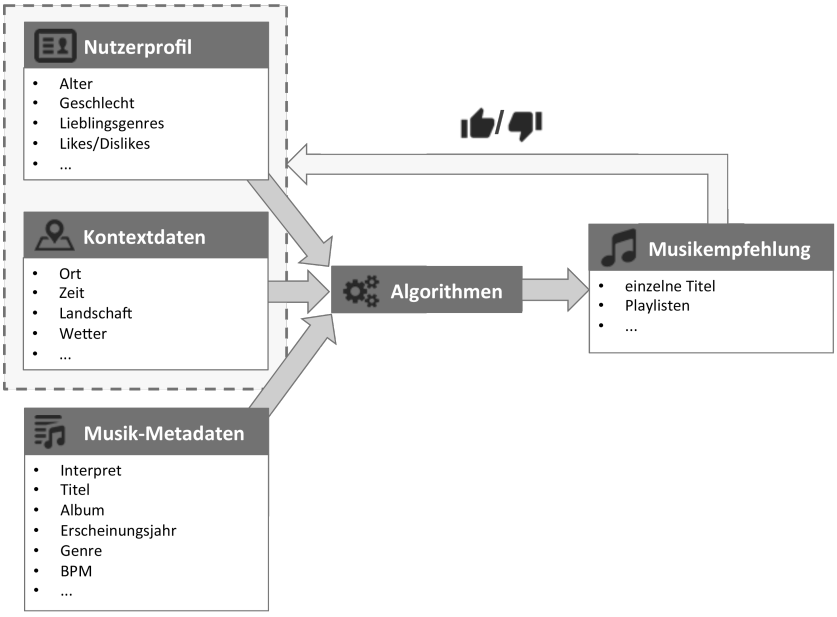


Abbildung 1: Funktionsweise kontextorientierter Musikempfehlungssysteme

Diese Systeme werden als kontextorientierte Musikempfehlungssysteme (engl. *Context-Aware Music Recommender Systems*, nachfolgend CAMRS genannt) bezeichnet [Ri12]. Das Interesse an diesen Systemen ist in den letzten Jahren besonders stark gestiegen. So

gibt es bereits Forschungsprojekte und daraus resultierende Prototypen, welche die Musikeinspielung anhand einzelner Kontextfaktoren, wie Ort, Zeit, körperliche Aktivitäten oder Wetter anpassen. [KR12, Le08, Ri12]

Trotz des steigenden Interesses im Bereich der CAMRS gibt es kaum Veröffentlichungen, die sich speziell mit dem Fahrtkontext und der Musikempfehlung im Fahrzeug beschäftigen. Im Forschungsprojekt *AmbiTune* werden Studien zur kontextorientierten Musikeinspielung im Fahrzeug durchgeführt und ein daraus resultierendes Empfehlungssystem entwickelt. Ein erster Prototyp einer Applikation und anschließende Realfahrtstudien sollen dazu beitragen, diese Forschungslücke zu schließen.

4 Studien zur Musikeinspielung im Fahrzeug

Ein Ziel des Projektes *AmbiTune* ist die Erhöhung der Fahrsicherheit durch weniger Benutzerinteraktion mit dem Bediensystem sowie durch gezielte Einspielung von Musik, die den Fahrer in komplexen Fahrsituationen weniger fordert. Zudem sollen Umweltfaktoren, die einen Einfluss auf die Musikeinspielung im Fahrzeug haben, identifiziert werden. Dazu wurde einerseits im Rahmen einer Fahrsimulatorstudie untersucht, wie sich unterschiedliche Musik auf die Fahrleistung auswirkt (siehe Kapitel 4.1). Andererseits wurde die Umfeldwahrnehmung des Fahrers anhand von Videostudien und Realfahrten untersucht, um relevante Kontextfaktoren zu identifizieren (siehe Kapitel 4.2).

4.1 Bewertung von Musik und Fahrsituationen

Um diese Einflüsse von Musik bei der Autofahrt näher analysieren zu können, wurden Probanden in einem Experiment verschiedener Musik ausgesetzt. Die Durchführung der Studie erfolgte mit Hilfe des Fahrzeugsimulators *Racer*[4]. Für das Experiment wurde eine Strecke mit einer Fahrzeit von ca. zehn Minuten gewählt. Während des Experiments wurde die Strecke zweimal befahren. In der ersten Fahrtrunde sollte einem vorausfahrenden Fahrzeug mit gleichbleibendem Sicherheitsabstand gefolgt werden. In der zweiten Runde hatten die Probanden freie Fahrt und sollten lediglich die Verkehrsregeln beachten (siehe Abbildung 2).

Während der Fahrzeit von jeweils zehn Minuten wurden den Probanden zehn unterschiedliche Lieder eingespielt. Die Lieder unterschieden sich im Tempo anhand der Beats per Minute (BPM). Jedes zweite Lied hatte ein hohes Tempo, während die andere Hälfte niedrige Tempi hatten. Die Lieder wurden in drei Blöcke eingeteilt, wobei ein

[4] Racer ist ein bereits von mehreren Forschern angewandter Open Source-Fahrsimulator, der sich durch seine hohe Realitätsnähe und umfangreiche Anpassungsmöglichkeiten auszeichnet. Zudem bietet das Programm die Möglichkeit, sämtliche Fahrdaten in Logfiles abzuspeichern (siehe www.racer.nl).

Abschnitt deutsche Lieder, ein zweiter englische und ein dritter instrumentale Lieder beinhaltete. Nach jedem Lied erfolgte die akustische Aufforderung zur Angabe der subjektiven Fahrbeeinflussung durch das Musikstück. Neben dieser eigenen Einschätzung wurden die Biodaten der Probanden in Form von Hautleitwert und Hauttemperatur erfasst. Zudem wurden die wesentlichen Fahrdaten, wie Geschwindigkeit und Distanz zum vorausfahrenden Fahrzeug durch das System aufgezeichnet. Die Musikeinspielung und Aufforderung der Bewertungsabgabe erfolgte über Kopfhörer, damit die Probanden keine Umgebungsgeräusche als Ablenkung wahrnehmen konnten.

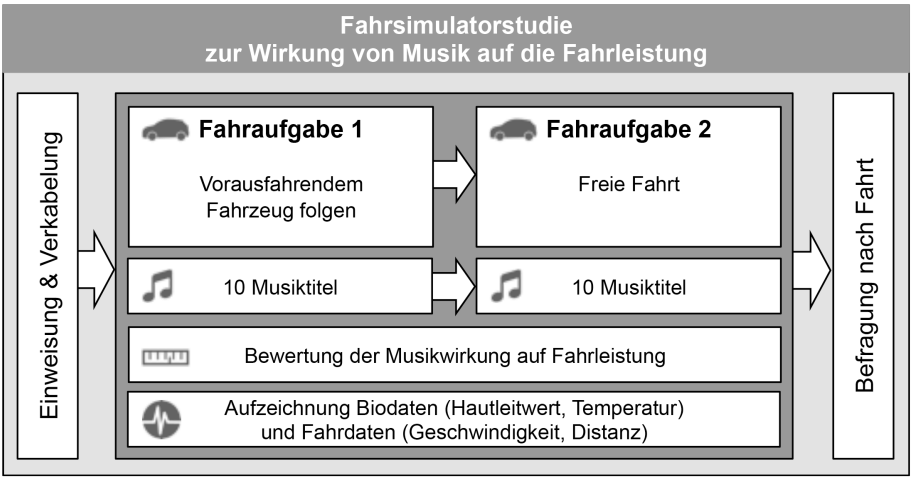


Abbildung 2: Aufbau der Fahrsimulatorstudie

Zwölf Frauen und 28 Männer mit einem Durchschnittsalter von 22,78 Jahren und einer Fahrerfahrung von mindestens zwei Jahren nahmen innerhalb einer Woche an dem Experiment teil. Die Teilnehmer wurden zufällig anhand ihrer Termine in drei Gruppen eingeordnet. Gruppe 1 und 2 bekamen Musik in unterschiedlicher Reihenfolge angeboten, wohingegen Gruppe 3 als Kontrollgruppe fungierte und ohne Musik fuhr.

Die Ergebnisse zeigen, dass das Tempo der Musikstücke die Fahrleistung beeinflusst. Generell zeigt sich, dass schnellere Lieder die Probanden stärker beansprucht haben als langsamere. Insbesondere die subjektive Empfindung der Belastung zeigte hier einen signifikanten stark positiven Zusammenhang zum Tempo der Musikstücke.[5] Auch bei den Körperfunktionen ließen sich tendenzielle Einflüsse von Musik auf Hautleitwert und Hauttemperatur nachweisen, die dieses Ergebnis stützen, jedoch nicht signifikant sind.

Ein Musikempfehlungssystem im Fahrzeug sollte demnach die aktuelle Verkehrslage bzw. die Belastung des Fahrers mit einbeziehen, um in solchen Fahrsituationen Musik mit hohem Tempo und starker Belastung zu vermeiden.

[5] Die Korrelation nach Pearson ist auf dem Niveau von $\alpha = 0{,}01$ (2-seitig) signifikant ($r = 0{,}803$; $p = 0{,}005$).

4.2 Wahrnehmung des Umfelds durch den Fahrer

Die folgende Studie betrachtet die Wahrnehmung des Umfelds durch den Fahrer unabhängig von der Musik. Dadurch soll erfasst werden, welche Umgebungsparameter in unterschiedlichen Fahrsituationen und unterschiedlichem Umfeld explizit durch den Fahrer erfasst werden. Dementsprechend kann eine Priorisierung der Berücksichtigung von Umgebungsparametern in unterschiedlichen Fahrsituationen erfolgen. In einer ersten Studie erfolgte die Analyse anhand von Videoszenen, bevor diese Ergebnisse mit Hilfe von Realfahrten validiert wurden.

Bei der durchgeführten Videostudie wurden 65 Testpersonen bezüglich ihrer persönlichen Wahrnehmung der Umgebung in acht unterschiedlichen Fahrtszenen befragt. Die Fahrtszenen wurden möglichst differenziert zwischen verschiedenen Straßenkategorien und Umgebungen ausgewählt. Dabei wurden die Probanden mit Kopfhörern vor einen Monitor gesetzt, um eine bestmögliche Wirkung der Szenen zu erreichen. Eine Szene besteht dabei aus einer 15-sekündigen Videoaufnahme durch die Frontscheibe eines fahrenden Autos und soll somit das Blickfeld des Fahrers widerspiegeln. Nach jeder dieser Szenen und einer Aufforderung über die Tonspur hatte die Testperson zehn Sekunden Zeit, die Umgebungswahrnehmung zu beschreiben und zusätzlich das Umfeld durch Ankreuzen auf einer Skala (sehr ländlich, eher ländlich, sowohl als auch, eher städtisch, sehr städtisch) einzuordnen (siehe Abbildung 3).

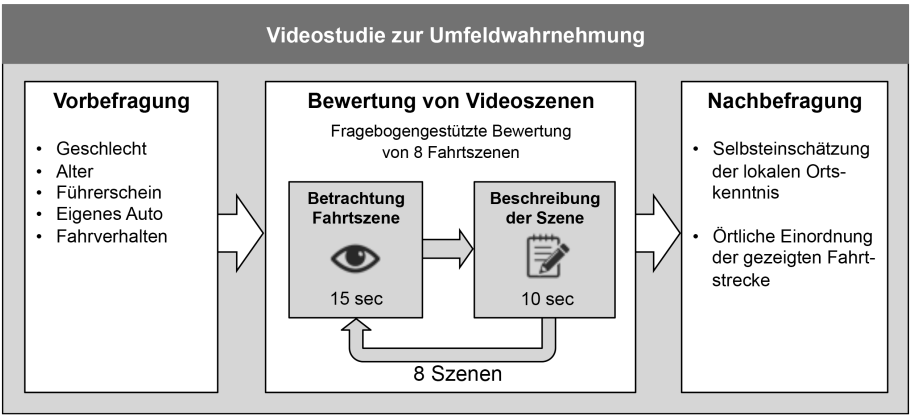


Abbildung 3: Aufbau der Videostudie

Es nahmen 17 Frauen (26,2 %) und 48 Männer (73,8 %) mit einem Durchschnittsalter von 23,42 Jahren an der Vorstudie teil. Einen Führerschein besaßen 62 Testpersonen (95,4 %). Zudem gaben 29 Personen (44,6 %) an, ein eigenes Auto zu besitzen. Die Ergebnisse zeigen grundlegend eine Dreiteilung der Umgebungswahrnehmung. Die Autobahn stellt dabei eine eigene Kategorie dar, die hauptsächlich durch die Fahrbahn und die Geschwindigkeit geprägt ist. Zudem ist eine Unterteilung in städtisches und ländliches Umfeld möglich. Das ländliche Umfeld ist eher durch gleichmäßige ländliche

Umgebung geprägt, während die Stadt durch vielfältige Umgebung, Menschen und Verkehr geprägt ist. Gerade im städtischen Bereich war das Spektrum der Umfeldbeschreibung durch die Probanden wesentlich größer. Es zeigte sich hier jedoch eine Tendenz in Richtung einer erhöhten Fahrtaufmerksamkeit durch die Angabe von Merkmalen wie Unübersichtlichkeit und Hektik. Zudem gab es Bereiche, die sowohl ländliche als auch städtische Merkmale aufwiesen.

In der anschließenden Fahrstudie fuhren fünf Probanden (2 weiblich, 3 männlich, Durchschnittsalter 24,2 Jahre) eine Strecke mit einer Gesamtlänge von 42 Kilometern und einer ungefähren Fahrzeit von 38 Minuten. Auch bei dieser Studie variierten sowohl die Straßenkategorie als auch das Umfeld der Strecke. Während der Fahrt selbst erfolgte kaum Interaktion mit den Probanden. Die Probanden wurden per Navigationsgerät entlang der vordefinierten Strecke geleitet und sollten während der Fahrt Umfeldänderungen angeben. Zudem wurden die Probanden an vordefinierten Testpunkten direkt zum wahrgenommenen Umfeld und zur kognitiven Belastung befragt. Während der Fahrt wurde das Blickfeld durch die Frontscheibe per Videokamera aufgezeichnet. Außerdem erfolgte eine Aufzeichnung der Angaben der Personen durch einen Audiorekorder. Weiterhin wurden die GPS-Positionen aufgezeichnet, um später eine Verbindung der jeweiligen Aussagen mit dem Ort herstellen zu können.

Die Ergebnisse bestätigten die Erkenntnisse aus der Videostudie. Auch hier wurde von den Probanden vor allem der städtische Bereich weitestgehend als Bereich mit hoher kognitiver Belastung wahrgenommen. Zudem zeigte sich auch hier, dass bei städtischen und ländlichen Gebieten die Umgebung stärker vom Fahrer wahrgenommen wird, was auf die Autobahn nicht zutrifft.

Ein Musikempfehlungssystem im Fahrzeug sollte somit zudem die Art des Umfelds der Strecke (z.B. städtisch oder ländlich) sowie die Straßenkategorie (z.B. Bundesstraße oder Autobahn) mit einbeziehen. Diese beiden Kontextparameter können mit weiteren Parametern angereichert werden und ermöglichen somit eine bessere Anpassung der Musik. Anhand der Straße lassen sich beispielsweise die Anzahl der Fahrspuren, das Geschwindigkeitslimit oder auch das Streckenprofil bestimmen.

5 Prototyp zur kontextorientierten Musikeinspielung

Anhand der Erkenntnisse aus den Studien wurde ein kontextorientiertes Musikempfehlungssystem für das Fahrzeug entwickelt, welches zudem die Nutzerbedürfnisse berücksichtigt. Ein erster Prototyp der AmbiTune-Applikation läuft einerseits als mobile Applikation auf einem Android-basierten Smartphone für den Einsatz im Fahrzeug und nutzt den integrierten GPS-Sensor zur Bestimmung des Ortes. Andererseits existiert ein Simulator als Webanwendung, der vor allem zum Testen und zur Veranschaulichung des Empfehlungssystems genutzt werden kann.

Neben der Ortsangabe wird der Vorhersagealgorithmus eines vorherigen Prototyps der Forschergruppe zur Bestimmung des Informationshorizonts verwendet. Dieser Prototyp beschreibt eine Vorausschau auf den bevorstehenden Streckenverlauf sowohl in geografischer als auch in zeitlicher Hinsicht. Er bietet somit die Grundlage zur Ermittlung weiterer relevanter Kontextparameter wie Zielort, Fahrtdauer oder Fahrtmission. Navigationssysteme bieten grundsätzlich die Möglichkeit, anhand des eingegebenen Ziels und der daraus errechneten Route den Informationshorizont zu generieren und weiteren Anwendungen zur Verfügung zu stellen. Jedoch finden Fahrten zu bekannten Zielen tendenziell ohne Navigationshilfe statt. Lediglich 6 Prozent der Autofahrer nutzen täglich ein Navigationsgerät [St11b]. Somit stehen diese notwendigen Kontextinformationen auf Alltagsfahrten nicht zur Verfügung und verhindern dadurch eine automatisierte Nutzung vieler Kontextparameter. Der Vorhersagealgorithmus nutzt das historische Fahrverhalten, um den Informationshorizont zu bestimmen. Weitere Informationen können [HZR13] entnommen werden.

Der Prozess der Musikempfehlung hin zu einer fahrt- und situationsbezogenen Playlist unterteilt sich in drei wesentliche Schritte (siehe Abbildung 4).

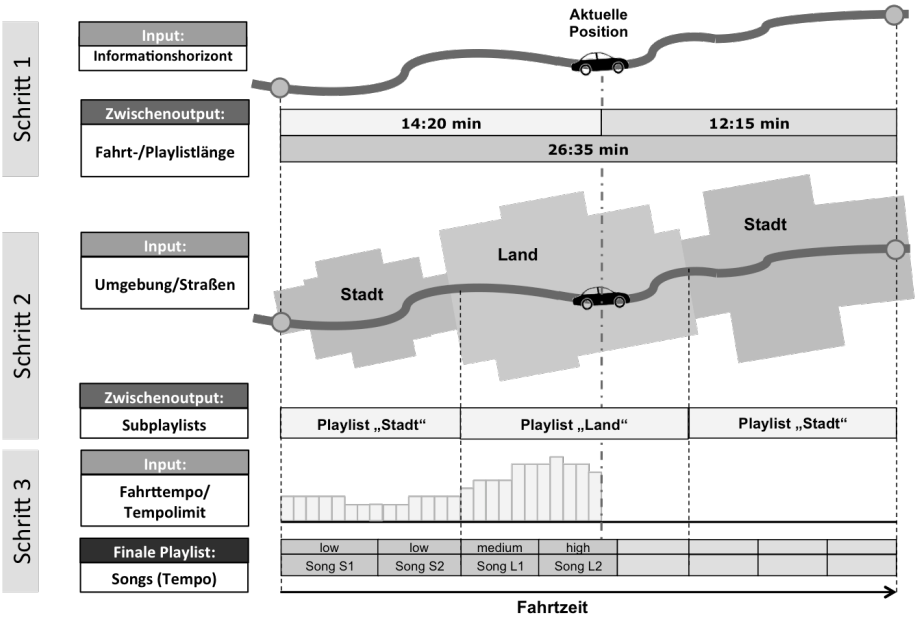


Abbildung 4: Eigenes Konzept zur Generierung einer Playlist

Im ersten Schritt dient der Informationshorizont als erster Input zur Bestimmung der ungefähren Länge der Playlist für die aktuelle Fahrt. In Schritt 2 können basierend auf den Positionsdaten des Informationshorizonts konsekutiv Anfragen an die

OpenStreetMap (OSM)[6]-Datenbank gestellt werden. Diese Anfragen liefern XML-Code zurück, welcher sämtliche Kartendaten für die aktuelle Position enthält. Anhand dieses Codes ist es möglich, bestimmte Parameter zu filtern und zu aggregieren. So erlauben Strecken- und Umgebungsdaten die Unterteilung der Playlist in unterschiedliche Subplaylists. Weiterhin kann anhand der konsekutiv aufgezeichneten GPS-Daten beispielsweise die Fahrtgeschwindigkeit ermittelt werden. Durch die Fahrtgeschwindigkeit in Bezug zum aktuellen Geschwindigkeitslimit und weiteren relevanten Kontextparametern können unterschiedliche Fahrsituationen ermittelt werden. Diese Fahrsituationen unterscheiden sich in der emotionalen Wirkung als auch in der Anforderung an den Fahrer. Die einzelnen Musikstücke aus der lokalen Musiksammlung des Nutzers können somit im dritten Schritt entsprechend unterschiedlichster Kontextparameter (hier Geschwindigkeit) automatisiert aus der Subplaylist ausgewählt werden.

Aktuell beinhaltet und verwendet *AmbiTune* zwölf unterschiedliche Kontextparameter, die je nach Fahrsituation mehr oder weniger relevant sind und daher priorisiert werden. Die relevanten Kontext- und Nutzerdaten werden in Kombination mit den vorhandenen Musikmetadaten genutzt, um die einzelnen Songs zur Einspielung auszuwählen.

Die Applikation bietet in einer ersten Version eine einfache Oberfläche, die dem Nutzer kaum Interaktion ermöglichen soll (siehe Abbildung 5). Ein einfacher Wechselschalter bietet die Möglichkeit, zwischen der musikbezogenen Sicht und der Kartensicht zu wechseln. In der musikbezogenen Sicht werden die aktuelle Playlist und das Genre angezeigt. Unterhalb dieser Angaben werden einige fahrtbezogene Daten angegeben.

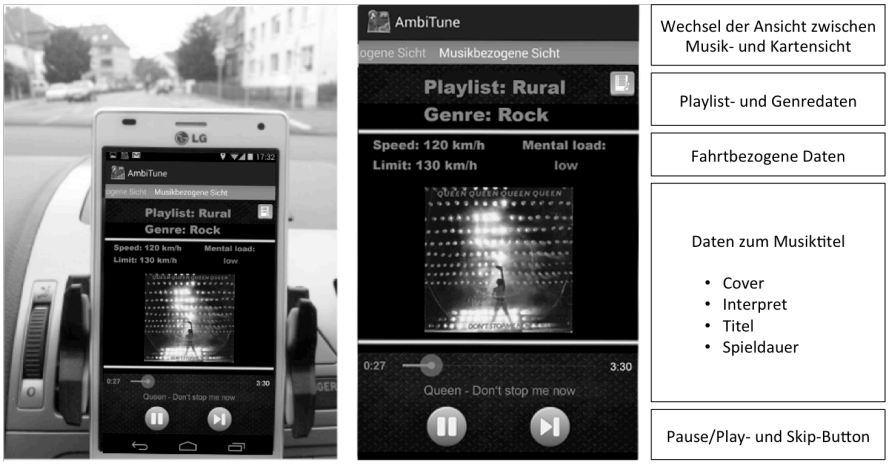


Abbildung 5: Oberfläche der Applikation auf einem Smartphone im Fahrzeug

[6] OSM ist ein kollaboratives Open-Source-Projekt zur Erstellung einer editierbaren Weltkarte und bietet detaillierte Informationen zu Straßen (z.B. Straßenkategorie, Fahrspuren, Tempolimit, Ampeln,) und zur Umgebung (z.B. Flächennutzung, Bebauung).

Hierbei handelt es sich um die aktuelle Geschwindigkeit des Fahrzeugs, das aktuelle Geschwindigkeitslimit der Strecke sowie die interpretierte kognitive Belastung des Fahrers. Neben diesen Kontextdaten werden die Daten des aktuell gespielten Musiktitels angegeben. Unterhalb des Songcovers erfolgt die Angabe der aktuellen Spieldauer des Songs sowie die Angabe des Interpreten und Titels. In der unteren Leiste der Applikation befinden sich lediglich zwei Buttons mit der Play- und Pausefunktion sowie der Möglichkeit, zum nächsten Song der aktuell generierten Playlist für die Fahrsituation zu springen.

6 Zusammenfassung und Ausblick

Die vorliegende Arbeit zeigt auf, dass der Bereich der CAMRS im Fahrzeug ein bisher wenig betrachteter Forschungsbereich ist, welcher jedoch durch steigende Sensorik und Konnektivität von Fahrzeugen zukünftig an Relevanz gewinnen wird. In dem Forschungsbereich besteht eine Forschungslücke zwischen der Untersuchung des kontextbezogenen Hörverhaltens im Fahrzeug und der Entwicklung von Prototypen. Das Forschungsprojekt *AmbiTune* unterstützt durch Studien und einer darauf aufbauenden Prototypentwicklung die Forschung in diesem Bereich und zeigt auf, wie eine automatische und kontextbezogene Anpassung der eigenen Musik im Fahrzeug ermöglicht werden kann. Hierzu kombiniert der eigene Ansatz primäre Sensordaten vom Smartphone und gefilterte sekundäre Kontextfaktoren aus dem Internet mit dem Nutzerprofil. Zudem dient der Prototyp als erster Ansatz für Realfahrtexperimente, wodurch eine frühzeitige Einbindung des Nutzers in den Entwicklungsprozess ermöglicht wird. Eine situationsbezogene Bewertung der Empfehlung durch den Nutzer kann als Kontrollinstanz und Evaluationsmethode genutzt werden. Neben der kontextbezogenen Auswahl von Musik soll perspektivisch auch die kontextbezogene Lautstärkenregulierung betrachtet und integriert werden.

Zukünftig ist die Anbindung der Applikation an eine Onlinemusikbibliothek wie Spotify oder Deezer geplant. Hierdurch soll eine Verbindung von einem kollaborativen und einem kontextorientierten Musikempfehlungssystem geschaffen werden. Dementsprechend können Musikempfehlungen generiert werden, die nicht nur auf ähnlichen Nutzerprofilen basieren, sondern auch von anderen Nutzern in ähnlichen (Fahr-)Situationen gehört wurden. Wird ein kontextbezogen empfohlener Song mit der aktuellen Fahrsituation gespeichert, kann einem anderen Nutzer mit ähnlichem Nutzerprofil der gleiche Song angeboten werden, wenn er in eine ähnliche Fahrsituation kommt.

Zudem ist in Anbetracht des erhöhten Interesses der führenden Entwickler mobiler Betriebssysteme[7] und der Standardisierung[8] des Datenaustauschs zwischen

[7] Hier sind vor allem Apples *CarPlay (*www.apple.com/ios/carplay) und Googles *Android Auto* (www.android.com/auto) zu nennen.
[8] Siehe *Open Automotive Alliance* (www.openautoalliance.net).

Applikationen und dem Fahrzeug eine Integration in das Fahrzeugsystem bzw. eine Anpassung an entsprechende Systeme entscheidend.

Literaturverzeichnis

[Ba10] Baltrunas, L. et.al.: Best Usage Context Prediction for Music Tracks. In: Proceedings of the 2nd Workshop on Context Aware Recommender Systems, 2010.

[BKR13] Braunhofer, M.; Kaminskas, M.; Ricci, F.: Location-aware music recommendation. In: International Journal of Multimedia Information Retrieval. 2 (1), S. 31–44, 2013.

[Br01] Brodsky, W.: The effects of music tempo on simulated driving performance and vehicular control. In: Transportation Research Part F: Traffic Psychology and Behaviour. 4 (4), S. 219–241, 2001.

[CM10] Cassidy, G.; Macdonald, R. (2010): The effects of music on time perception and performance of a driving game. In: Scandinavian journal of psychology. 51 (6), S. 455–464, 2010.

[CS14] CSC, Auto Infotainment: 61 Prozent der deutschen Autofahrer fühlen sich abgelenkt., http://www.csc.com/de/press_releases/113265-auto_infotainment_61_prozent_der_deutschen_autofahrer_fuehlen_sich_abgelenkt, Stand: 18.01.2015.

[De11] DeNora, T.: Music in Everyday Life. Cambridge University Press, 2011.

[DW07] Dibben, N.; Williamson, V.: An exploratory survey of in-vehicle music listening. In: Psychology of Music. (35), S. 571–589, 2007.

[Ga12] Garzon, S.: Intelligent In-Car-Infotainment Systems: A Contextual Personalized Approach. In: 2012 8th International Conference on Intelligent Environments (IE). S. 315–318, 2012.

[Ga14] Gartner, Gartner Identifies the Top 10 Strategic Technology Trends for 2015. Abgerufen am 10.10.2014 von http://www.gartner.com/newsroom/id/2867917, 2015.

[Ge14] Gerla, M. et.al.: Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds. In: 2014 IEEE World Forum on Internet of Things (WF-IoT). S. 241–246, 2014.

[HZR13] Helmholz, P.; Ziesmann, E.; Robra-Bissantz, S. (2013): Context-awareness in the Car: Prediction, Evaluation and Usage of Route Trajectories. In: Proceedings of the 8th International Conference on Design Science at the Intersection of Physical and Virtual Design. Berlin, Heidelberg, S. 412–419, 2013.

[KR12] Kaminskas, M.; Ricci, F.: Contextual music information retrieval and recommendation: State of the art and challenges. In: Computer Science Review. 6 (2–3), S. 89–119, 2012.

[Le08] Lehtiniemi, A.: Evaluating SuperMusic: streaming context-aware mobile music service. In: Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology. New York, S. 314–321, 2008.

[MTL11] Mitsopoulos-Rubens, E.; Trotter, M.; Lenné, M.: Effects on driving performance of interacting with an in-vehicle music player: A comparison of three interface layout concepts for information presentation. In: Applied Ergonomics. 42 (4), S. 583–591, 2011.

[NH99] North, A.; Hargreaves, D.: Music and driving game performance. In: Scandinavian Journal of Psychology. 40 (4), S. 285–292, 1999.

[NHH04] North, A.; Hargreaves, D.; Hargreaves, J.: Uses of Music in Everyday Life. In: Music Perception: An Interdisciplinary Journal. 22 (1), 2004.

[Ra13] Radiozentrale.de, ma 2013 Radio II, http://www.radiozentrale.de/studien-und-daten/radionutzung/ma-2013-radio-ii/, Stand 13.01.2015.

[Ri12] Ricci, F.: Context-aware Music Recommender Systems: Workshop Keynote. In: Proceedings of the 21st International Conference Companion on World Wide Web. New York, S. 865–866, 2012.

[RKR07] Reichwald, R.; Krcmar, H.; Reindl, S.: Mobile Dienste im Auto der Zukunft. 1., Aufl. Eul Verlag, Lohmar, 2007.

[SH08] Seppänen, J.; Huopaniemi, J.: Interactive and context-aware mobile music experiences. In: Proceedings of the 11th Int. Conference on Digital Audio Effects. Espoo, 2008.

[St11a] Statista, Situationen der Radionutzung bis 2011, http://de.statista.com/statistik/daten/studie/208895/umfrage/situationen-der-radionutzung-jahresvergleich/, Stand: 12.01.2015.

[St11b] Statista: Nutzungshäufigkeit Pkw-Navigationssysteme, http://de.statista.com/statistik/daten/studie/180738/umfrage/nutzungshaeufigkeit-von-pkw-navigationssystemen-in-deutschland/, Stand: 20.11.2014.

[St12] Statista, Radionutzung - Wege im Jahr 2012 | Umfrage, http://de.statista.com/statistik/daten/studie/208887/umfrage/wege-der-radionutzung/, Stand: 28.11.2014.

[St14] Statista, Die größten Musik-Streaming Seiten 2014, http://de.statista.com/statistik/daten/studie/221795/umfrage/die-groessten-kostenpflichtigen-musik-streaming-dienste/, Stand: 24.01.2015.

[Th11] Thomes, T.: An Economic Analysis of Online Streaming: How the Music Industry Can Generate Revenues from Cloud Computing. Discussion Paper Nr. 11-039, Rochester, 2011.

[USE12] Ünal, A.; Steg, L.; Epstude, K.: The influence of music on mental effort and driving performance. In: Accident; analysis and prevention. 48 , S. 271–278, 2012.

[Va13] Van der Zwaag, M. et.al.: Using music to change mood while driving. In: Ergonomics. 56 (10), S. 1504–1514, 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Automotive MILS

Jean-Pierre Seifert[1] Michael Peter[2]

Abstract: Ohne Elektronik, die im Laufe der letzten 30 Jahre ein unverzichtbarer Bestandteil im Automobilbau geworden ist, wäre eine Vielzahl von Innovationen wie ABS, EPS etc. undenkbar gewesen. Dieser Trend wird sich aller Wahrscheinlichkeit nach auf absehbare Zeit mit Fahrerassistenz- und Infotainmentsystemen fortsetzen, wenn nicht sogar verstärken. Vor dem Hintergrund der stark wachsenden Systemkomplexität ist die bisherige Annahme, dass alle Komponenten als vertrauenswürdig betrachtet werden können, fragwürdig. Die daraus resultierenden Fragen nach der Systemsicherheit und -verlässlichkeit können ohne geeignete Gegenmaßnahmen zu einer erheblichen Innovationsverlangsamung führen.

Angesichts der sich abzeichnenden Anforderungen, mehr Funktionen auf einer Plattform zu integrieren, schlagen wir den Einsatz von MILS (multiple independent levels of security), einer bewährten Sicherheitssystemarchitektur, im automobilen Umfeld vor. Wir skizzieren anhand von drei Beispielen wie mit Hilfe von MILS Sicherheits- und Verlässlichkeitsrisiken begegnet werden kann: Absicherung von Kommunikation in offenen Netzen, Einsatz von Standardbetriebssystemen (Linux) und Softwareaktualisierung (FOTA).

[1] Technische Universität Berlin, jpseifert@sec.t-labs.tu-berlin.de
[2] Technische Universität Berlin, peter@sec.t-labs.tu-berlin.de

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Static Value Range Analysis for Matlab/Simulink-Models

Christian Dernehl[1] Norman Hansen[1] Thomas Gerlitz[1] Stefan Kowalewski[1]

Abstract: In this paper we present a static value range analysis of signals within function block diagrams developed with Matlab/Simulink. We analyse signals with respect to their data type and compute an approximation of the possible value range of each signal represented by a set of intervals. In addition the analysis reports potential problems within the model, like occurrences of NaN (Not-a-Number) and divisions by zero or infinity. We show the applicability of our analysis on a viscosity model from Matlab Central and a model from the set of automotive example models that is contained within Matlab/Simulink.

Keywords: Automotive software engineering, static analysis, Matlab/Simulink

1 Introduction

Model-based software development is used extensively within the automotive domain with one of the tools used being Matlab/Simulink. Simulink allows the creation of control models in the form of functional block diagrams, in which the data flow is expressed as a sequence of blocks connected by lines. In functional block diagrams, blocks model a function between several inputs to multiple outputs. The combination of simple functions, such as standard arithmetic, allows the modeling of complex systems as a composition of blocks from a basic block set. This design is supported by building hierarchies, to allow the reuse of existing systems and, at the same time, abstract from simple blocks. After modeling the control model, code can directly be generated from it, which is then compiled and loaded onto the target platform.

During the modeling process of a control model, errors might occur, which might eventually result in a total failure of the control system. To prevent such errors, models are tested and simulated exhaustively against a design specification. Especially in the context of safety critical systems, the behavior of a control system has to be verified to operate within specified design ranges to conform standards like the ISO26262 [IS11]. Nevertheless testing might, depending on the specification, not detect all errors and flaws within a model. Applying formal analysis in early development stages is a possible extension of the established verification process.

In this paper we present an analysis which computes an approximation of the state space of a control system modeled with Matlab/Simulink by abstracting the state space into interval sets, which are propagated along the path from inputs to outputs within the control model. Using this technique, the developer specifies sets of input intervals, as required by the ISO26262, and gets a safe approximation set of potential output intervals, which can

[1] RWTH Aachen, Informatik 11 - Embedded Software, surname@embedded.rwth-aachen.de

be used to verify the previously specified design ranges. The specification with intervals provides a simple and formal way to express the input constraints. For instance, consider an external input from an accelerometer, whose input ranges by specification are $[-4g, 4g]$, which can be read in most cases of the data sheet. Next to the resulting proposals for design ranges of each subsystem, as denoted in the ISO 26262, our analysis provides proof of the absence of values becoming NaN, division by zero/infinity and over- and underflows of signals.

Unlike abstract interpretation of the resulting C code, our analysis interprets the model. In many cases, the code generated from the model includes checks for array out of bounds and overflows, which cannot be detected by a code analysis, because the generated code handles these cases implicitly. Consider a vector, from which an element is selected using the Selector block[5], which allows to specify the index externally. In this case, code generators may include a check for the index and if the index is out of bounds, zero is returned. This mechanism proves hard to detect for static code analysis, since the array index is actually never out of bound. However, an activation of the safety procedure is hidden to the engineer.

The paper is structured as follows. In the next Section, related work in the field of value-range approximation of functional block diagrams and other domains is presented. Section 3 contains the methodology of our approach, which covers the used abstraction domain of interval sets, the set of supported functions by the analysis and how loops are approximated. The analysis is evaluated in Section 4 using a model of an open-loop and a closed-loop system. Finally, potential extensions of the approach are shown and the paper is concluded in Section 5.

2 Related Work

Thinking of the automatic verification methods for software systems, various approaches have been developed for different domains. Value set analysis, which provides a value set for each signal or variable in the program can be used to identify potential errors such as division by zero, underflows and overflows of the program variables and dead code in the program. An analysis of the resulting code from the model has been carried out to prove the absence of such errors [Si08]. Opposed to an analysis on the generated code, the Design Verifier[6] operates directly on the model by using the polyspace static analysis engine.

Depending on the configurations, the generated code of the Simulink model might be catching division by zero errors, which are still occurring in the production code, while the tester is unaware of them. Consider the case, where there is an automatic check if the divisor is zero and then the result of the function would be just 0, which is better than a runtime exception, but hides the problem. Instead our approach uses the semantics of the complete Simulink model and interprets each block individually.

[5] See `http://mathworks.com/help/simulink/slref/selector.html` .

[6] The Design Verifier is a Tool for Simulink which allows similar analysis of the model as the presented approach. `http://www.mathworks.com/products/sldesignverifier/index.html?s_tid=gn_loc_drop`

Further approaches to verify such block diagrams are performed by transforming them into an intermediate representation which is easier to analyze. For instance, converting the discrete part of Simulink to Lustre allows the utilization of existing analysis tools for synchronous languages [Tr05, Ca03]. In [CD06] a subset of the available block set within the modeling tool Simulink is transformed into an equivalent representation of the timed interval calculus, which can be used to prove formal requirements and estimate the expected value ranges of specific signals. A similar approach is presented in [CM09] where an abstract interpretation of Simulink models is presented to compute potential rounding errors when using float calculations, by computing an over approximation using Taylor series.

More exhaustive approaches include property provers in their analysis to prove the absence of bugs. In their work [RG14], Reicherdt and Glesner convert Simulink models to Boogie, which uses the Microsoft Z3 sat modulo theory (SMT) prover to prove constraints. These constraints are generated from the blocks and types in the Simulink model. Similar work has previously been done by Bauch et. al [BHB14], where models are converted to boolean formulas, using the bit vector representation.

All of the presented solutions do either an analysis of the generated code, which is very useful to check if the generated code has now flaws, or, as the latter approaches, focus on an exhaustive analysis and precise or rather numerical related computations. Therefore, different transformations into other formal representation formats are used. A general difference of our work compared to those approaches concerns the set of blocks and functionality which is supported by the described transformations and the consecutive analyses. For instance, there are limitations regarding the support of multi dimensional signals [RG14] and blocks for structural operations on such [CD06, RG14], restricting the supproted blocks to per definition *safe* blocks [CD06] or relating to the scalability of the approach for potentially very large models [BHB14].

3 Fault Detection Methods for Control Models

After having given an overview of related approaches, the focus is on our proposal. Since our analysis is based on value sets, intervals are used to represents those sets. By using sets of intervals, a more precise representation of values can be given. For each block in the model, we define a semantic with respect to finite interval sets with intervals for floating point valued signals. A finite interval set $S_I = \{I_0, \ldots, I_n\}$ contains a finite number of non-overlapping intervals $I_i, 0 \leq i \leq n$, where the interval might be either open, half-open or closed. First we define interval sets and the underlying types we use. Afterwards, the semantic for a selection of blocks, which are supported by our analysis, is presented.

3.1 Intervals and Interval Sets

Interval arithmetic is well known and widely used in different application areas [Ja12], [HJVE01], [BKS12]. However, compared to the common usage of interval arithmetic,

we check additionally faults for the IEEE 754 float. This is because many application scenarios are based on a hardware floating point unit, which eases the development process, because conversions to fixed point calculations is not required. Therefore, the symbols $-\infty$, ∞ and *nan* are also possible values in the considered interval domain. The arithmetic for intervals containing these symbols has already been implemented and defined [Wa98]. Assume the set of valid IEEE 754 floats to be $\mathbb{R}^{IEEE}$, then, the following holds $x+\infty=\infty$ (for $x\neq-\infty$), $x-\infty=-\infty$ (for $x\neq\infty$), $x*\infty=\infty$ (for $x\neq 0$), $x\circ nan=nan$ (for $\circ\in\{+,-,*,\div\}$). Note, that $[nan,nan]$ is the interval representation of the IEEE-symbol *nan*, [ ] represents the empty interval and $[-\infty,\infty]$ is a valid non-empty interval.

Interval Set Arithmetic Consider an interval set S_I and any well-defined function $f:\mathbb{R}^{2\times n}\mapsto\mathbb{R}^2$, as in [NJC99], mapping n input intervals to an output interval. For interval sets $S_0,\ldots S_n$, we define $f_S(S_0,\ldots,S_n)$ as the union of the output interval sets, where f is applied to each combination of the input intervals.

$$f_S(S_0,\ldots,S_n)=\bigcup_{I_0\in S_0}\ldots\bigcup_{I_n\in S_n}f(I_0,\ldots,I_n) \tag{1}$$

For functions with n arguments and interval sets of cardinality m the resulting set may have m^n elements. However, the cardinality could shrink when applying functions to an interval set S, because two regions might overlap or have a common border. For instance, $\{[2;2],[4;4]\}+\{[-1;1]\}=\{[1;5]\}$. Nevertheless, our analysis limits the size of intervals of an interval set to 20, merging intervals together when the limit is exceeded, since the consecutive operations and over-/underflows might increase the number of intervals exponentially. Still, that limit is not reached in our applications. This limit can be configured to arbitrary positive integer values in case larger or smaller interval sets, leading to potentially smaller or bigger overapproximations, are desired by the user. Since the function f, serves in most cases as an operator, we call the process of applying f or f_S respectively an *operation*.

Typing We extend the concept of interval sets to typed interval sets by associating a datatype to every interval set. The currently supported datatypes are signed integer (int8, int16, int32), unsigned integer (uint8, uint16, uint32), floating point (float, double) and logical (boolean) types.

The datatype of an interval set is a property which restricts the set of values representable by an interval set. This means that an interval set of type int8 can only represent signed integer values which can be stored within 8 bits. To assure that every interval set S_i of datatype $\Delta\in\{boolean,int8,int16,int32,uint8,uint16,uint32,float,double\}$ always complies with its datatype restrictions, a type-restriction procedure ρ_Δ is applied after every operation on the interval set. The restriction operations ρ_Δ can be categorized into three categories $BOOL=\{\rho_{boolean}\}$, $FLOAT=\{\rho_{float},\rho_{double}\}$ and $INT=\{\rho_{int8},\rho_{int16},\rho_{int32},\rho_{uint8},\rho_{uint16},\rho_{uint32}\}$.

The boolean nature of an interval set is created by applying $\rho_{boolean}(S_i)$ which yields an interval set containing 0 (false) in case $0\in S_i$ and 1 (true) in case $\exists v\in S_i, v\neq 0$, while

the restriction methods of the category $FLOAT$ restrict the values of the interval set to single precision (float) or, respectively, to double precision (double) floating point values according to the IEEE 754 standard. In contrast to operations of categories $BOOL$ and $FLOAT$, all operations $\rho_{int} \in INT$ have to restrict the value range of integer values with potential under- and overflows. Due to the use of interval sets, under- and overflows can be represented without overapproximations. For instance, for interval sets yields $\rho_{int8}(\{[100;128]\}) = \{[-128;-128],[100;127]\}$, whereas $\rho_{int8}([100;128]) = [-128;127]$ yields a large overapproximation for intervals.

Both interval sets have to be of the same datatype, so that binary operations can be applied. This is no restriction compared to operations performed in Simulink or other programming languages since those perform implicit type-castings. To be able to perform these casts explicitly we provide correspondent operations for all supported datatypes which correspond to the application of the restriction method $\rho_{targettype}$. Since the interval sets are internally represented using the most expressive type, most casts can be performed using the type system of the underlying architecture. However, the cast from floating to integer types can be performed using different rounding modes. We focus on those rounding modes from [IE08] which are supported (but named differently) by Simulink, too:

- *round up*: Rounds towards $+\infty$.
- *round half to even*: Rounds to the nearest representable number. If these is ambiguous, rounds to the nearest even number.
- *round down*: Round towards $-\infty$.
- *round half up*: Rounds to the nearest representable number. If these is ambiguous, applies *round up* mode.
- *round half away from zero*: Rounds to the nearest representable number. If these is ambiguous, rounds positive values as *round up* and negative values as *round down*.
- *truncate*: Round towards 0.

Dimensions Signals in Simulink models can be of arbitrary dimension. Accordingly our domain of interval sets is extended to multidimensional matrices of interval sets where each matrix element is an interval set. Thus, operations on signals are operations on matrices of interval sets in our domain. Matrix operations in Simulink are either structural operations, e.g. concatenation along a dimension, or can be mapped to operations on the interval sets. For instance, the matrix multiplication of two matrices can be written as multiple expressions consisting of multiplications and additions. For simplicity of the explanations, we describe directly the operations on interval sets which is identical to the correspondent operation on matrices of exactly one element.

3.2 Operations

The following list represents a subset of function blocks, which are supported by our tool. We do not describe in this paper the standard arithmetic for intervals and the continuous

non-linear math operations, which are supported, since those have already been defined for intervals and can be mapped to interval sets as described.

- Standard arithmetic: $+, -, *, /$
- Continuous non-linear math: exp, log, trigonometry
- Relational Operators: $>, \geq, <, \leq, =, \neq$
- Logical Operators: $\neg$, $\wedge$, $\vee$
- Decisions, Switches
- Interpolation/Lookup Tables

Relational Operators First we describe, how relations are applied to interval sets. Note, that the result of a relational operator on intervals is generally not binary, but a three valued logic $\{0, 1, \perp\}$, where $\perp$ represents that the result is unknown. For instance, the outcome of the relational operation $[0,5] > [-5,3]$ is unknown. With interval sets, $\perp$ can be defined as $\{[0,0],[1,1]\}$. Let S_A and S_B be interval sets and $\succ_\perp \in \{>, \geq, <, \leq, =, \neq\}$ be a relation. Following the definition of Equation (1) $S_A \succ_\perp S_B$ is defined as $\cup_{I_A \in S_A} \cup_{I_B \in S_B} I_A \succ_\perp I_B$. In this case, the union operator might merge $[0,0]$ with $[1,1]$, representing an unknown result, although each interval relation $\succ_\perp$ yields a defined result. For example $\{[-1,-1],[1,1]\} > [0,2]$ yields for each element of the left hand side a known result, while the union of all is unknown. Nevertheless, the use of interval sets provides a more precise representation when casting boolean expressions to floating point values, e.g. as a multiplier, so that it is clear that 0.5 is not in the set.

Logical Operators For the logical operators $f_{\mathbb{B}} : \mathbb{B}^n \mapsto \mathbb{B}$, we define a new function $f_S(u_0 \neq 0, \ldots, u_n \neq 0)$ for *real* valued input interval sets $u_0, \ldots, u_n$, where 0 means $\{[0,0]\}$ in this case. Note, that $\neq$ has been previously defined on interval sets. This inequality check to zero corresponds to the standard C casting operations. The resulting output is generally calculated by applying Equation (1) on the interval sets, where the corresponding f for n inputs, and, thus, n intervals must be specified. Nevertheless, since the domain is limited to true and false, the algorithm can terminate to compute the output if two valid $x_0 \in u_0, \ldots, x_n \in u_n$ are found, making $f_{\mathbb{B}}$ become true and false.

Decisions Some blocks, like the switch block, pass a certain input through them depending on a condition c, i.e. input a is passed through if condition c is valid, otherwise input b goes through, denoted as $\text{ite}(c,a,b)$. The condition receives one parameter in form of a value computed by the model which is used to formulate the actual condition c as an relation containing the parameter value, where the result of the logical relation is an interval set as described before. If the result of the relation is $\{[0,0]\}$ or $\{[1,1]\}$, then a or respectively b is passed through the block. Additionally, a warning is issued to the user, indicating that there is a dead path in his model. However, if the result of the relation is $\{[0,0],[1,1]\}$, $a \cup b$ is the result, since the condition cannot be uniquely evaluated. Remember, due to the use of *NaN* values, the entire path of the condition has to be evaluated, and a shortcut by

evaluating for example $sin(u) > 1$ without thinking of u cannot be made, since u might be NaN.

Lookup Tables Often the behavior of specific parts in control models can be estimated with measurements, which are then fed into the model during the development process. These measurement points then realize a new function within the model. Thus, for $n \geq 2$ points $x_0, \ldots, x_n \in \mathbb{R}$ with $x_i < x_j$ for $(i < j)$ and $y_0, \ldots, y_n \in \mathbb{R}$ measurements, a function $LT(x_i) = y_i$ is defined in the model. Hence, the x_i and y_i are predefined and during execution another input x is provided, so that the output y is set to y_i if $x = x_i$. If x is in between two values x_i, x_{i+1}, then the result is either interpolated or clipped to the last y_i, which is specified by the engineer. Defining the associated interval set function for lookup tables is a two step process, in which firstly helper functions LT_C and LT_B are defined, where LT_C represents the selection of interpolation or clipping by the engineer and LT_B is a special expression handling inputs that are out of bounds, i.e. $x < x_0$ or $x > x_n$. Secondly using the ite operator, the lookup can be recursively defined. With regard to linear interpolation, the function $LT_C(I_x, y_i, y_{i+1})$ becomes

$$LT_C(I_x, y_i, y_{i+1}) \quad = \quad \left[\min_{x \in I_x} \frac{x}{x_{i+1} - x_i}(y_{i+1} - y_i), \max_{x \in I_x} \frac{x}{x_{i+1} - x_i}(y_{i+1} - y_i) \right] \tag{2}$$

where $I_x = [\underline{x}, \overline{x}]$ is an interval with $x_i \leq \underline{x}$ and $x_{i+1} < \overline{x}$ is a linear interpolation, which is extended to intervals. Regarding a flat behavior, the function becomes constant $LT_C(I_x, y_i, y_{i+1}) = [y_i, y_i]$. With respect to LT_B, this function checks if the lower bound is reached and uses otherwise the upper bound $LT_B(I_x) = \text{ite}(x < x_0, y_0, y_n)$. In the next step, ite is applied to check whether the input is between x_i and x_{i+1} for all valid i. If $x > x_n$, y_n is returned, yielding the following expression

$$LT_S(X) = \bigcup_{I_x \in X} \left(\bigcirc_{0 \leq i \leq n-1} \text{ite} \quad x_i \leq I_x \wedge I_x < x_{i+1} \quad LT_C(I_x, y_i, y_{i+1}) \right) LT_B(I_x) \tag{3}$$

where $\circ$ is the function application and X is the input interval set. Using currying, the bounded recursive application of the ite operator specifies a piecewise function, representing the behavior of the lookup table. For example, if $n = 3$, then the expression becomes

$$LT_S(X) = \bigcup_{I_x \in X} \text{ite}(x_0 \leq x \wedge x < x_1, y_0, \text{ite}(x_1 \leq x \wedge x < x_2, y_1, \text{ite}(x < x_0, y_0, y_n))) \tag{4}$$

where the last stacked ite becomes $LT_B(I_X) = \text{ite}(x < x_0, y_0, y_n)$. Note that LT_B only returns y_n in case $x > x_n$ since all other cases are covered by the previous ites.

3.3 Loops and Widening

In the previous section, we described how different functional blocks can be interpreted with interval sets. Considering open-loop systems whose outputs do not depend on a state

of the system, this approach is sufficient because for each block on the path the output intervals can be estimated. For closed-loop systems with hierarchical or nested loops, an approximation has to be performed, to compute the result set in finite time. Since the model is executed using a fixed step time Δt, the output interval sets of the model can be calculated for a fixed time horizon T within $T\Delta t^{-1}$ steps. However, the analysis targets not only the consideration of a given time horizon, as done by simulations, but that the system runs indefinitely. For Simulink models with loops corresponds the computation of a time step to the computation of a loop iteration for all loops in the model. For instance, the computation of all signal values for time step 0 corresponds to the computation of the first iteration of every loop in the model.

Widening is a well known abstraction procedure of values and time to analyze value ranges of variables in combination with loops. Therefore, during the analysis of the loop, a widening operator on the abstract domain overapproximates the computed reachable values [CC77]. Applying the widening operator, if chosen properly, will finally ensure that the analysis reaches a fixed-point regarding the reachable values of the loop and finally the entire system. When representing loops as a function $f(u,y)$, a fixed-point is a point y, for which $f(u,y) = y$ holds. This means, that further considered loop iterations (applications of $f(u,y)$) will not change the already computed reachable values y. We apply the widening approach as presented in [CC77] for the static value range analysis of program code to Simulink models. This has the particularity that a model might contain implicit loops introduced by Simulink-blocks of specific types. We call these loops implicit, since they are no loops which can be seen by looking on the connections between blocks in Simulink. Implicit loops are introduced by blocks which generate a non-constant outgoing signal based on no or constant input signals, e.g. Integrator-, Ramp- or Step-blocks. Figure 1 shows a simple example for a system with an implicit loop. Although, it is an open-loop system, we apply widening for the output of the integrator to ensure a correct overapproximation of the reachable values and the existence of a fixed-point for the model on the analysis domain. In case widening would not be used, the output of the Integrator-block would change for any considered timestep, a fixed-point would never be reached and the analysis would not terminate.

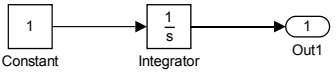


Fig. 1: Open loop system with implicit Simulink loop

A second particularity regarding widening and the associated detection of a fixed-point for Simulink models is the "'delay along paths"' (DAP). Assume a system containing a loop with a Delay-block. Such a block delays its incoming signal for some timesteps. The model shown in Figure 2, for instance, matches this description. The delay block delays the incoming signal for three timesteps, which means that its outgoing signal is by default zero for those timesteps. If widening is directly applied for the second loop iteration (second timestep), no overapproximation would be made and a fixpoint detected, because a further loop iteration does not change any computed reachable values. However, considering further iteration, the delay changes its outgoing values. Such fixpoint detection problems are avoided by calculating the maximal DAP, which is the maximal number of

timesteps (respectively loop iterations) a signal value needs to influence the value at the end of the path in the model, and apply widening on the abstract domain only in case enough timesteps have been considered. Because of the manifold ways to create dependencies between loops in Simulink models, we compute the maximal delay of the entire model as overapproximation of the DAP for single paths and use this value instead. Moreover, we take the maximal delay of the model into account for fixed-point detection to ensure, that no further iteration of any loop may change the computed reachable values. In case of the model from Figure 2, the maximal delay of the model is three and equal to the maximal delay of the path from the Constant- to the Outport-block.

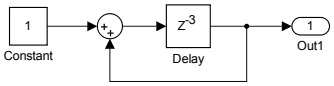


Fig. 2: Explicit loop with DAP from Constant to Out1 of three

4 Evaluation

The aim of the evaluation of our proposal is to show the feasibility on the one hand, while discussing drawbacks on the other hand. For the feasibility, the analysis has been executed on a computer with a 2.6 GHz 64 bit Intel i5 4-core CPU with 8GB memory running Windows 7 and 64 bit Java 8. With respect to the data, the evaluation is carried out on two models, both configured to a fixed step size of 0.01. The first model is a viscosity model and illustrated in Figure 3.

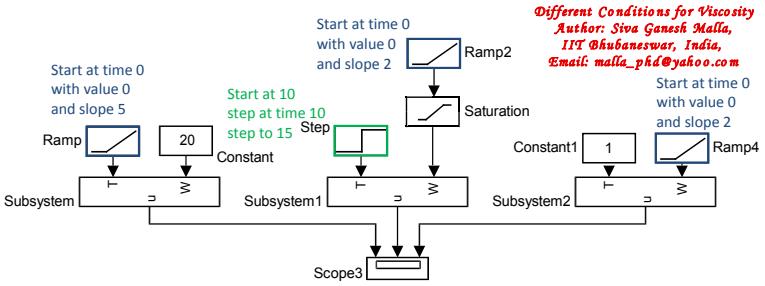


Fig. 3: Viscosity Model (taken from [Ma12]) with annotations

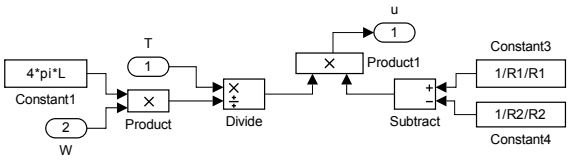


Fig. 4: Viscosity Model Subsystem (taken from [Ma12])

The internals of the models subsystems are identical and shown on Figure 4. However, the values for L, R1 and R2 are different from subsystem to subsystem, since they are configured using masks. Table 1 shows the values for each subsystem (rounded for better readability). The specific parameters for the Ramp-, Step- and Saturation- blocks are annotated in Figure 3.

	Subsystem	Subsystem1	Subsystem2
L	14e-3	14	14
R1	10.5e-3	10.5	10.5
R2	12e-3	12	12

Tab. 1: Mask parameters of the subsystems (Viscosity model)

The analysis of the viscosity model takes about 13.782 seconds and determines the reachable output values for the subsystems as given in Table 2. Potential problems, e.g. the occurrence of *NaN* as a value, of the model are indicated through warning messages. For instance, the analysis indicates a potential division by zero at the Divide-block in Subsystem1 and Subsystem2 and a potential division by infinity at the Divide-block in Subsystem2. This is accurate considering that the Ramp-blocks produce signals starting at zero and that the system runs for an infinite amount of time.

	Subsystem	Subsystem1	Subsystem2
Reachable values at outport	$\{[0;\infty]\}$	$\{[6.0478E-6;0.0060418],[\infty;\infty]\}$	$\{[0;6.0417E-4],[\infty;\infty]\}$

Tab. 2: Results of the value range analysis (Viscosity model)

The second model we chose for evaluation purpose is the "'Automotive Suspension"' [TM] model from the Simulink example models. This model contains 44 basic Simulink blocks of 10 different types (Sum, Gain, Step, Constant, Integrator, Mux, Demux, Subsystem, Outport, Inport). In contrary to the viscosity model, it contains feedback-loops that depend on each other. We replaced the continuous integrators of the model with discrete integrators, which use the forward euler method. That is because the support for different continuous integration solvers is not yet completely finished. Furthermore, we switched the "'acceleration due to gravity"' block, which was originally a Constant-block, to an Inport-block. This allows to vary the gravity depending on the considered location. Since our static analysis of the model operates on interval sets, we are able to analyze the model for arbitrary user specified value ranges of the gravity. We chose as input for the gravity the interval sets $A = [-9.81;-9.81]$, $B = [-9.80665;-3.71]$ and $C = [-9.82;-9.80]$. Those sets represent the commonly used value for gravity in our region (A), the range of gravity values above the zero level on earth [TT08] and mars[7] (B) and an arbitrary, but similar to A chosen range (C). Since we know that the model converges to a fixpoint for the constant -9.81, we perform the analysis without use of a widening operator, which yields a simulation of the model on the interval set domain terminating in case a fixpoint is reached. Regarding the results of the simulation with interval set input A we could determine that our results match exactly (using 64 bit double precision) the values computed by Simulink for every signal of the model at any considered time. This gives us a hint on the numerical compliance of our implementation with Simulink.

In general, we use widening to ensure the existence of a fixpoint for systems with loops [CC77]. Table 3 shows our results[8] using a simple widening procedure for the discussed

[7] `http://nssdc.gsfc.nasa.gov/planetary/factsheet/marsfact.html`

[8] We show the results as intervals instead of interval sets and rounded for better readability

Input Interval Set	Widening	THETA (Integrator out)	Z (Integrator out)	Time elapse(s)
A	OFF	$[-0.01098; 0.00862]$	$[-0.15372; 0.0]$	39.779
A	ON	$[-\infty; \infty]$	$[-\infty; \infty]$	31.772
B	ON	$[-\infty; \infty]$	$[-\infty; \infty]$	17.192
C	ON	$[-\infty; \infty]$	$[-\infty; \infty]$	17.570

Tab. 3: Results of the value range analysis (Automotive Suspension model)

inputs. The results indicate, that widening (if enabled) is applied before a fixpoint could be determined and thus a huge overapproximation is achieved. This overapproximation leads to warnings with false positive character indicating potential problems where none can occur. For instance, the possible occurrence of the value *NaN* as a result of the operation $[-\infty; \infty] + [-\infty; \infty]$ is indicated due to the overapproximation. Our future work focuses on approaches to reduce this amount of overapproximation for loops using alternative widening approaches specific to the static analysis of Simulink models, as other additional analysis domains.

5 Conclusion

This paper presents a formal approach to perform static value range analysis of functional block diagrams, in particular Matlab/Simulink, before generating code. The approach allows the direct detection of modeling flaws during the modeling phase, early in the development process. Based on user-provided value ranges for inputs, as demanded by ISO26262 [IS11], a safe value range of outputs is computed. Therefore, the abstract domain of interval sets, a more precise representation than the widely applied intervals, is used. We show, that the application of a type concept to this domain can cope very precisely with numeric issues, e.g. rounding modes, and type related over- and underflows. The evaluation of the approach shows that it is suited to detect potential errors as divisions by zero, infinity or occurrences of NaN. In the future, we will evaluate the value range analysis on more realistic examples. However, the use of interval sets can lead to similar large overapproximations as intervals in combination with widening for loops of the model. Thus, the approach is not sufficient to describe the behavior of a dynamic system, but indicates potential modeling flaws. The next step is to represent the relations between signals of the model, for example $u_0(t) > u_1(t)$, and the rates of the signals. With this information more precise conclusions, e.g. reduction of potential overapproximations, about the system can be drawn.

References

[BHB14] Bauch, Petr; Havel, Vojtech; Barnat, Jirí: Accelerating temporal verification of Simulink diagrams using satisfiability modulo theories. Software Quality Journal, pp. 1–27, 2014.

[BKS12] Biallas, Sebastian; Kowalewski, Stefan; Schlich, Bastian: Range and Value-Set Analysis for Programmable Logic Controllers. In: Proceedings of the 11th International Workshop on Discrete Event Systems. IFAC, Guadalajara, Mexico, pp. 378–383, 2012.

[Ca03] Caspi, Paul; Curic, Adrian; Maignan, Aude; Sofronis, Christos; Tripakis, Stavros; Niebert, Peter: From Simulink to SCADE/Lustre to TTA: A Layered Approach for Distributed Embedded Applications. In: Proceedings of the 2003 ACM SIGPLAN Conference on Language, Compiler, and Tool for Embedded Systems. LCTES '03, ACM, New York, NY, USA, pp. 153–162, 2003.

[CC77] Cousot, P.; Cousot, R.: Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints. In: Conference Record of the Fourth Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages. ACM Press, New York, NY, Los Angeles, California, pp. 238–252, 1977.

[CD06] Chen, Chunqing; Dong, JinSong: Applying Timed Interval Calculus to Simulink Diagrams. In (Liu, Zhiming; He, Jifeng, eds): Formal Methods and Software Engineering, volume 4260 of Lecture Notes in Computer Science, pp. 74–93. Springer Berlin Heidelberg, 2006.

[CM09] Chapoutot, A.; Martel, M.: Abstract Simulation: A Static Analysis of Simulink Models. In: Embedded Software and Systems, 2009. ICESS '09. International Conference on. pp. 83–92, May 2009.

[HJVE01] Hickey, T.; Ju, Q.; Van Emden, M. H.: Interval Arithmetic: From Principles to Implementation. J. ACM, 48(5):1038–1068, September 2001.

[IE08] IEEE: IEEE Standard for Floating-Point Arithmetic. IEEE Std 754-2008, pp. 1–70, Aug 2008.

[IS11] ISO: ISO 26262-6 - Road vehicles - Functional safety - Part 6 Product development software level. Technical report, Geneva, Switzerland, 2011.

[Ja12] Jaulin, L.; Kieffer, M.; Didrit, O.; Walter, E.: Applied Interval Analysis: With Examples in Parameter and State Estimation, Robust Control and Robotics. Springer London, 2012.

[Ma12] Malla, Siva: Calculation of Viscocity. 2012. `http://de.mathworks.com/matlabcentral/fileexchange/39413-calculation-of-viscocity`.

[NJC99] Nedialkov, N.S.; Jackson1, K.R.; Corliss, G.F.: Validated solutions of initial value problems for ordinary differential equations. Applied Mathematics and Computation, 105(1):21 – 68, 1999.

[RG14] Reicherdt, Robert; Glesner, Sabine: Formal Verification of Discrete-Time MATLAB/Simulink Models Using Boogie. In (Giannakopoulou, Dimitra; Salaün, Gwen, eds): Software Engineering and Formal Methods, volume 8702 of Lecture Notes in Computer Science, pp. 190–204. Springer International Publishing, 2014.

[Si08] Simon, A.: Value-Range Analysis of C Programs. Springer, August 2008.

[TM] The MathWorks, Inc.: Automotive Suspension. `http://de.mathworks.com/help/simulink/examples/automotive-suspension.html?prodcode=SL`.

[Tr05] Tripakis, Stavros; Sofronis, Christos; Caspi, Paul; Curic, Adrian: Translating Discrete-time Simulink to Lustre. ACM Trans. Embed. Comput. Syst., 4(4):779–818, November 2005.

[TT08] Thompson, Ambler; Taylor, Barry N.: Guide for the Use of the International System of Units (SI). National Institute of Standards and Technology Special Publication 811 (2008 Edition). National Institute of Standards and Technology / U.S. Department of Commerce, 2008.

[Wa98] Walster, G. William: The Extended Real Interval System. Technical report, 1998.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

SLRefactor: Ein Refactoring-Ansatz für Simulink-Modelle

Quang Minh Tran [1] Jonas Winkler [2] Christian Dziobek [3]

Abstract: Bei der Funktionsmodellierung ist die Veränderung und Erweiterung der Struktur eines Modells eine häufig durchgeführte Aktivität. Während es bereits Refactoring-Ansätze für textuelle Programmiersprachen wie Java, C# usw. gibt, fehlt ein vergleichbarer, integraler und durchgehender Ansatz für Simulink-Modelle. Wir haben einen automatisierten Refactoring-Ansatz (im Folgenden SLRefactor-Ansatz genannt) für Simulink-Modelle erfolgreich entwickelt, der in einem Zeitraum von ca. zwei Jahren in der Serienentwicklung bei der Daimler AG erprobt und eingesetzt wurde. In diesem Beitrag wird der SLRefactor-Ansatz anhand eines ausführlichen Beispiels erläutert und es wird über die Erfahrungen beim produktiven Einsatz des Ansatzes und über die dabei gewonnenen Erkenntnisse berichtet.

Keywords: Modellbasierte Entwicklung, Simulink, Refactoring, Transformation

1 Einführung

Zur Beherrschung der kontinuierlich zunehmenden Komplexität wird Fahrzeugsoftware in der Automobilindustrie modellbasiert entwickelt. Hierzu hat sich MATLAB/Simulink [Ma14] als Defacto-Standard bei vielen großen Automobilherstellen, wie zum Beispiel bei der Daimler AG [DRW12], etabliert. Mit MATLAB/Simulink werden Steuer-/Regelalgorithmen in Fahrzeugsoftware als Simulink-Blockdiagramme modelliert. Simulink bietet viele Vorteile. Diese bestehen insbesondere in Hinblick auf übersichtliche grafische Darstellung und hohe Freiheitsgrade beim Entwurf. Der Einsatz der modellbasierten Entwicklung hat zur Folge, dass die Erstellung von Funktionsmodellen ein wesentlicher Schritt bei der Entwicklung von Fahrzeugfunktionen in Kraftfahrzeugen geworden ist. Aus den Modellen wird durch den Einsatz von Codegeneratoren die Funktionssoftware für die Steuergeräte im Fahrzeug generiert. Daher ist sowohl die Qualität der Modelle in Bezug auf Struktur und Anordnung der Blöcke, als auch die Benennung von Blöcken, Signalen und Ports von großer Bedeutung.

Bei der Erstellung, Überarbeitung und Weiterentwicklung eines Modells werden typischerweise zahlreiche strukturelle Änderungen durchgeführt, um bestehende Modellteile mit Blick auf verbesserte Struktur neu zu organisieren, überflüssige Modellteile zu entfernen oder das Modell strukturell auf Erweiterungen vorzubereiten. Mit den aktuellen Mitteln des Simulink-Editors (aktuelle Version: R2014b) lassen sich diese Ziele nur mit einem großen manuellen Aufwand erreichen, weil dieser lediglich elementare Bearbeitungsoperationen bereitstellt. Der Editor bietet Operationen zum Platzieren, Verbinden

[1] Daimler Center for Automotive IT Innovations, Ernst-Reuter-Platz 7, Berlin, quang.tranminh@dcaiti.com
[2] Daimler Center for Automotive IT Innovations, Ernst-Reuter-Platz 7, Berlin, jonas.winkler@dcaiti.com
[3] Daimler AG, 71059 Sindelfingen, christian.dziobek@daimler.com

und Gruppieren von Blöcken zu einem Subsystem (*Create Subsystem*) und Auflösen eines Subsystems (*Expanse Subsystem*). Allerdings fehlen höherwertige funktionsorientierte und kontextabhängige Operationen, die die Modellerstellung und Modellstrukturierung als auch die Änderung bestehender Struktur effizient unterstützen. Aus diesem Grund muss derzeit beispielsweise eine Verbindung zweier Blöcke, die miteinander Informationen austauschen, vom Anwender durch sequentielle Anwendung vieler elementarer Bearbeitungsoperationen (z.B. Hinzufügen von *Inport-*/*Outport*-Blöcken und Linien) erstellt werden. Es fehlt eine Operation, die diese Verbindung in einem Schritt erstellt, gegebenenfalls auch über mehrere Hierarchieebenen hinweg.

In den Veröffentlichungen von den Autoren [QD13] und [TWD13] wurden die ersten Ideen eines Ansatzes zum Refactoring von Simulink-Modellen vorgestellt. Seitdem wird der Ansatz weiterentwickelt und in der Vor- und Serienentwicklung bei der Daimler AG erprobt. In diesem Beitrag wird der aktuelle Stand dieses Ansatzes beschrieben, der *SLRefactor* genannt wird. Dazu werden zunächst kurz die Basiselemente von Simulink-Modellen vorgestellt.

2 Grundlegende Simulink-Modellierungselemente

In diesem Abschnitt werden die wichtigsten Simulink-Modellierungselemente kurz beschrieben. Abb. 1 zeigt dazu ein einfaches Simulink-Modell.

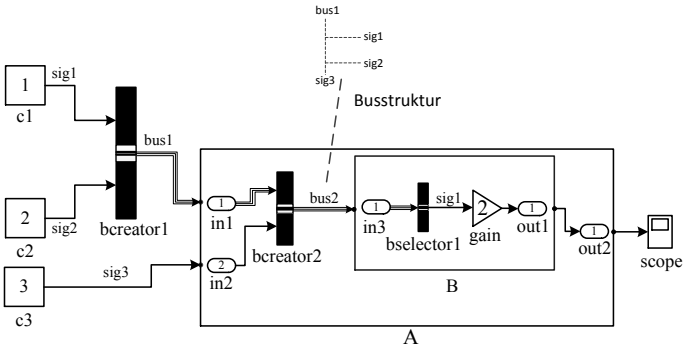


Abb. 1: Beispiel eines Simulink-Blockdiagramms

Ein Simulink-Blockdiagramm ist ein Signalflussgraph und besteht aus Blöcken und Linien. Blöcke führen in der Regel Berechnungen durch, beispielsweise Addition und Integration. Die Eingabe- und Ausgabeschnittstelle eines Blocks wird durch seine Eingangs- und Ausgangsports definiert. Blöcke werden durch Linien verbunden. Über diese Linien werden Daten in Form von Signalen zwischen den Blöcken übertragen. Ein Subsystem ist ein besonderer Block, der andere Blöcke einschließlich anderer Subsysteme enthalten kann. Subsysteme ermöglichen es, Simulink-Modelle hierarchisch zu strukturieren. In Abb. 1 enthält das Subsystem `A` die *Inport*-Blöcke `in1` und `in2`, den *BusCreator*-Block `bcreator2`, das Subsystem `B` und den *Outport*-Block `out2`. Ein Subsystem hat ebenfalls Eingangs- und Ausgangsports, die seine Eingabe- und Ausgabeschnittstelle beschreiben.

Mehrere Signale können durch *BusCreator*-Blöcke zu Bussen zusammengefasst werden. Busse können andere Busse enthalten und weisen damit eine baumartige Signalstruktur auf. Ein Signal in einem Bus kann durch einen *BusSelector*-Block ausgewählt werden.

In Abb. 1 werden die Signale `sig1` und `sig2` durch den *BusCreator*-Block `bcreator1` zu dem Bus `bus1` zusammengefasst. Dieser Bus wird wiederum zusammen mit dem Signal `sig3` durch den *BusCreator*-Block `bcreator2` zu dem größeren Bus `bus2` gruppiert. Die Struktur des Busses `bus2` wird in der Abbildung veranschaulicht. Das Signal `sig1` wird mit Hilfe des *BusSelector*-Blocks aus `bus2` selektiert.

3 Problemstellung

Wie eingangs erwähnt ist die Erstellung und Weiterentwicklung von Simulink-Modellen ein sehr aufwändiger und zeitintensiver Entwicklungsschritt. Simulink-Modelle mit mehr als 20 000 Blöcken sind in der Praxis keine Seltenheit. Zur Beherrschung der hohen Komplexität wird ein Simulink-Modell typischerweise hierarchisch unter Verwendung von Subsystemen organisiert. Auf oberster Modellebene wird die Gesamtfunktionalität in mehrere Module unterteilt. Die durch Subsysteme realisierten Module kapseln zusammenhängende Funktionsanteile und tauschen untereinander Informationen aus. Da zwischen Modulen auf oberster Ebene sehr viele Signale übertragen werden, werden häufig Signale zu Bussen zusammengefasst. Innerhalb eines Moduls wird die Funktionsaufteilung hierarchisch weiter verfeinert. Abb. 2 zeigt die oberste Ebene eines solchen Modells. In diesem Modell kommunizieren die Module `F1`, `F2` und `F3` untereinander über Busse. Eingangs- und Ausgangssignale sind ebenfalls Busse.

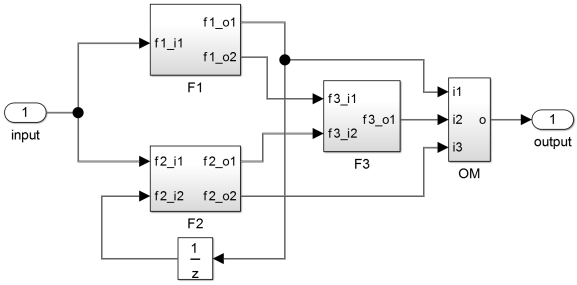


Abb. 2: Eine typische Modularchitektur des Modells

Strukturänderung von Modellen ist eine häufige Aktivität bei der Modellierung. Insbesondere sollen häufig die Schnittstellen zwischen den Modulen modelliert werden. Als nächstes sollen Module restrukturiert werden, wie zum Beispiel durch das Zerlegen eines Moduls oder Zusammenfassen mehrerer Module. Wenn das Modell eine gute Struktur erhält, sollen häufig bestimmte Funktionalitäten im Modell erweitert oder geändert werden. Dazu müssen unter anderem Signale von einem Modul zu einem anderen Modul weitergeleitet oder unnötige Signale entfernt werden. Bei jeder Strukturänderung muss darauf geachtet werden, die für das Modell vereinbarten Muster beizubehalten.

Um alle diese Aufgaben zu erledigen, fehlt allerdings integrierte Unterstützung für höherwertige Operationen in der heutigen Simulink-Umgebung. Vielmehr muss der Entwickler heute auf die elementaren Editierungsoperationen wie beispielsweise Hinzufügen von Blöcken und Linien zurückgreifen.

4 Katalog von Refactoring-Operationen für Simulink

Ausgehend von einer detaillierten Analyse von typischen Anwendungsfällen bei der Bearbeitung von Simulink-Modellen ist ein Katalog von Refactoring-Operationen für Simulink entstanden. Dieser Katalog ist an einen vergleichbaren Katalog für objektorientierte Programmiersprachen [Fo99] angelehnt[3]. Abb. 3 zeigt den Katalog in Form einer Mind-Map. Der Katalog enthält alle Refactoring-Operationen und gruppiert diese anhand der Art der am Simulink-Modell durchgeführten Änderung. Im Folgenden werden die Hauptkategorien dieses Katalogs vorgestellt.

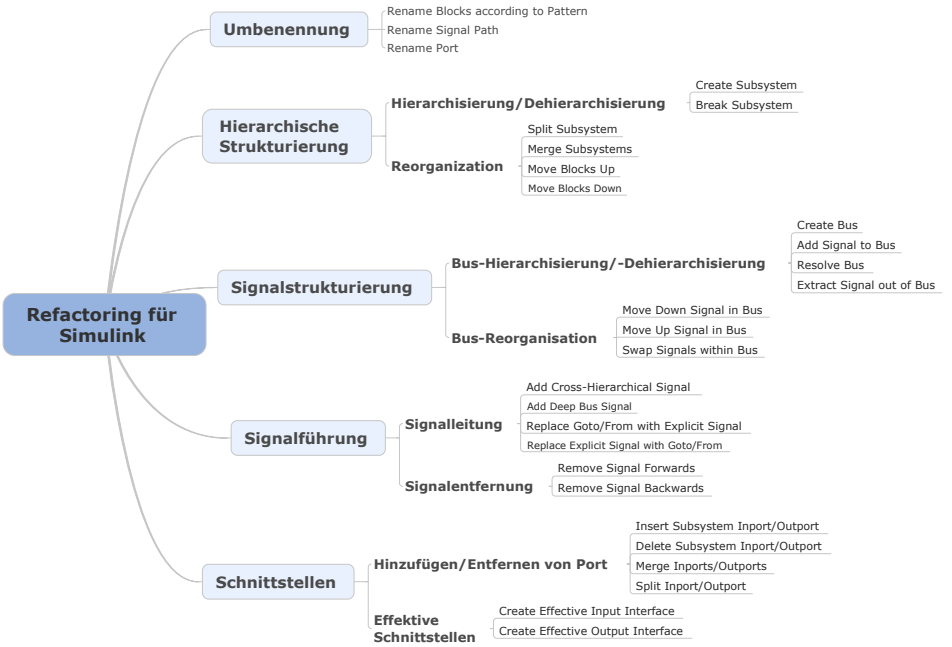


Abb. 3: Katalog von Refactoring-Operation für Simulink

Umbenennung: In einem Simulink-Modell sind die Namen von Blöcken, Subsystemen, Signalen und Ports ein wichtiges Beschreibungselement und beeinflussen unmittelbar die Verständlichkeit und Lesbarkeit des Modells. Der Simulink-Editor unterstützt bereits die Umbenennung einzelner Elemente. Allerdings sind derzeit komplexe Umbenennungen, wie zum Beispiel die Umbenennung aller Ports entlang eines Signalpfades oder die An-

[3] Martin Fowler pflegt auf seiner Webseite einen aktuellen Katalog von Refactoring-Operationen für objektorientierte Programmiersprachen. Der Katalog ist abrufbar unter der Adresse: http://refactoring.com/catalog/

wendung eines Namensmusters auf alle Elemente innerhalb eines Subsystems nicht möglich. Die Refactoring-Operationen in dieser Kategorie ermöglichen eine effiziente Umbenennung von Blöcken, Signalpfaden und Ports.

Hierarchische Strukturierung: Bei der Modellerstellung spielt die Partitionierung eines Modells in hierarchisch angeordnete Subsysteme sowie die Definition der Schnittstellen und die Festlegung des Informationsflusses zwischen den Subsystemen durch Signale eine große Rolle. Diese Kategorie umfasst Refactoring-Operationen zur Bearbeitung der Modellstruktur. *Create Subsystem* fasst Blöcke zu einem Subsystem zusammen, während die inverse Operation *Break Subsystem* ein Subsystem durch seinen Inhalt ersetzt. *Move Blocks Up* verschiebt Blöcke auf die nächst höhere Subsystemebene und *Move Blocks Down* verschiebt Blöcke auf die nächst niedrigere Subsystemebene. *Split Subsystem* teilt ein Subsystem in zwei kleinere Subsysteme. Die dazu inverse Operation *Merge Subsystems* fasst zwei Subsysteme zu einem einzigen Subsystem zusammen. Bei all diesen Operationen werden die Signalbeziehungen der verschobenen Blöcke beibehalten.

Signalstrukturierung: Häufig werden Signale zu Bussen zusammengefasst, um so logische Zusammenhänge zu definieren und damit einhergehend die visuelle Komplexität zu reduzieren. Die Refactoring-Operationen dieser Kategorie dienen der Bearbeitung hierarchischer Busstrukturen. *Create Bus* erzeugt einen Bus aus mehreren Signalen. *Add Signal to Bus* fügt ein Signal einem Bus hinzu. *Resolve Bus* löst einen Bus in seine einzelnen Signale auf. *Extract Signal out of Bus* überführt ein Signal eines Busses in ein elementares Signal. *Move Down Signal in Bus* und *Move Up Signal in Bus* erlauben eine Verschiebung eines Signals entlang einer Bushierarchie.

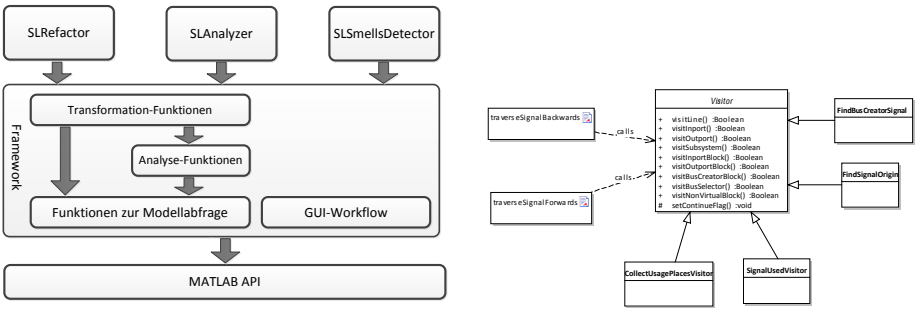
Signalführung: Für die Beschreibung von Algorithmen in Simulink müssen Signalinformationen über Subsysteme hinweg verteilt werden. Aus diesem Grund ist die Führung eines Signals über die Subsystemhierarchie hinweg eine sehr häufig durchgeführte Aufgabe bei der Modellerstellung. Die zu dieser Kategorie gehörenden Refactoring-Operationen dienen dazu, neue Signale über Subsystemgrenzen hinweg hinzuzufügen bzw. bestehende Signalpfade zu entfernen. *Add Cross Hierarchical Signal* erzeugt eine Signalverbindung zwischen zwei Ports über die Subsystemhierarchie hinweg, in dem Inport-/Outport-Blöcke und Linien auf dem Pfad zwischen den Ports hinzufügt werden. *Add Deep Bus Signal* führt ein Signal zu einem Ziel. Dabei wird das Signal durch Busse weitergeleitet. *Replace Goto/From with Explicit Signal* ersetzt korrespondierende *Goto/From*-Blöcke durch eine explizite Signalverbindung, die gegebenenfalls über die Subsystemgrenzen hinweg verlaufen kann. Hingegen ersetzt die inverse Operation *Replace Explicit Signal with Goto/From* eine Signalverbindung durch *Goto/From*-Blöcke. Bei *Remove Signal Backwards* wird ein Signalpfad bis zur Signalquelle oder bis zu einer vorgegebenen Hierarchieebene entfernt. Hingegen entfernt *Remove Signal Forwards* in Vorwärtsrichtung alle Signalpfade, die aus einem Signal hervorgehen.

Schnittstellen: Die Eingabe- und Ausgabeschnittstelle eines Subsystems wird durch die Eingangs- und Ausgangsports des Subsystems definiert. Bei der Erstellung und Bearbeitung eines Modells spielt die Definition der Subsystemschnittstellen zur Verdeutlichung des Informationsflusses eine wichtige Rolle. Die Refactoring-Operationen dieser Kategorie unterstützen den Entwickler daher bei Veränderung der Schnittstellen. *Insert Subsystem*

Inport/Outport fügt einen neuen Eingangs- bzw. Ausgangsport an einer beliebigen Position zu einem Subsystem hinzu. *Remove Subsystem Inport/Outport* entfernt einen Eingangs- bzw. Ausgangsport an einem bestimmten Index. Mit Hilfe von *Merge Ports* können Signale mehrerer Ports zusammengefasst und mit *Split Port* auf mehrere Ports verteilt werden. *Create Effective Input/Output Interface* erzeugt für ein Subsystem zusätzlich ein sogenanntes INMAP-bzw. OUTMAP-Subsystem. Für eine detailiertere Darstellung des Modellierungsmusters mit INMAP- und OUTMAP-Subsystemen sei der Leser auf [Ra02] verwiesen.

5 Werkzeugkonzept und Implementierung

Im Rahmen dieser Arbeit ist ein Refactoring-Framework für Simulink entwickelt worden, welches im Folgenden *SLRefactor-Framework* genannt wird. Das Framework wurde mithilfe der objektorientierten Sprache M-Script realisiert. Das SLRefactor-Framework definiert einen GUI-basierten Workflow in abstrakter Weise mit Hilfe des *Template Method*-Entwurfsmusters [Ga94]. Das Refactoring-Werkzeug *SLRefactor* benutzt das Framework und implementiert die Refactoring-Operationen des Katalogs.



(a) Architektur der Umsetzung (b) Besucher-Entwurfsmuster

Abb. 4: Umsetzung

Die in Abb. 4(a) veranschaulichte Schichtenarchitektur des SLRefactor-Frameworks ist durch das Eclipse-Refactoring-Framework [Wi06] inspiriert und besteht aus den folgenden Bausteinen:

Funktionen zur Modellabfrage von Modellen Das Framework stellt eine Sammlung höherwertiger Utility-Funktionen zum Abfragen von Modellen bereit.

Infrastruktur zur Analyse von Simulink-Modellen: Die meisten Refactoring-Operationen erfordern eine Analyse des Modells. Beispielsweise muss bei der Refactoring-Operation *Add Cross Hierarchical Signal* zum Erstellen eines ebenen-übergreifenden Signalflusses zwischen zwei Ports unterschiedlicher Ebenen berechnet werden, wie das Signal über die Hierarchieebenen hinweg geführt werden muss. Hierfür muss das *letzte gemeinsame Subsystem* beider Ports ermittelt werden.

Zusätzlich stellt die Infrastruktur Funktionen zur Verfolgung von Signalen bereit. Zur Signalverfolgung gehören Vorwärts- und Rückwärtsverfolgung eines Signals. Vorwärtsverfolgung wird genutzt, um die Blöcke zu ermitteln, an denen ein Signal verwendet wird. Die Analyse der Signalnutzung stellt eine wesentliche Voraussetzung für die Operationen *Remove Signal Forwards* und *Create Effective Input Interface* dar. Analog dazu dient die Rückwärtsverfolgung dazu, die Quelle eines Signals zu finden. Die Refactoring-Operation *Remove Signal Backwards* benutzt diese Analyse, um zu ermitteln, bis zu welchem Block ein Signal ohne Auswirkungen auf andere Modellteile entfernt werden kann. Des Weiteren lässt sich für ein in einem Bus geführtes Signal durch Rückwärtsverfolgung herausfinden, an welchem *BusCreator*-Block der Bus erzeugt wird. Diese Analyse ist unter anderem relevant, um die Refactoring-Operationen *Move Down/Up Signal in Bus* und *Swap Signals in Bus* umzusetzen.

In der Implementierung wird eine Traversierungsstrategie (Vorwärts- und Rückwärtsverfolgung) und eine konkrete Logik durch eine Variante des Besucher-Entwurfsmusters [Ga94] getrennt (siehe Abb. 4 (b)). Außerdem sind die Funktionen zur Vorwärts- und Rückwärtsverfolgung in der Lage, ein Signal durch Busse und über Subsystemgrenzen hinweg zu verfolgen.

Infrastruktur zur Transformation von Simulink-Modellen: Neben den Analyse-Funktionen stellt das SLRefactor-Framework mehrere Transformationsschritte bereit. Transformationsschritte sind Funktionen, die atomare Änderungen an Simulink-Modellen durchführen. Es wird zwischen elementaren und zusammengesetzten Transformationsschritten unterschieden. Ein elementarer Schritt modifiziert ein Simulink-Modell, ohne andere Schritte zu benutzen. Beispiele für elementare Schritte sind *addLine* zum Hinzufügen einer Linie zwischen Ports derselben Hierarchieebene und *addInport/addOutport* zum Hinzufügen eines *Inport-*/*Outport*-Blocks zu einem Subsystem.

Ein zusammengesetzter Schritt wird durch eine Abfolge elementarer oder zusammengesetzter Schritte definiert. Beispielsweise ist der Schritt *addCrossHierarchical* zum Hinzufügen eines ebenenübergreifenden Signals zwischen zwei Ports unterschiedlicher Hierarchieebenen ein aus aus *addLine*, *addInport* und *addOutport* bestehender zusammengesetzter Schritt.

SLRefactor-Werkzeug: Das *SLRefactor*-Werkzeug benutzt die Infrastruktur des Frameworks und implementiert die meisten Refactoring-Operationen aus dem Katalog in Abb. 3. Das Werkzeug wurde in den Simulink-Editor integriert. Die Refactoring-Operationen können wie in Abb. 5 gezeigt während der Modellierung über ein Kontext-Menü aufgerufen werden. Zur Verbesserung des Modell-Layouts nach einer Refactoring-Operation wird ein automatischer Layouting-Algorithmus für Simulink-Modelle [Kl12] verwendet.

SLAnalyzer und SLSmellsDetector: Neben *SLRefactor* sind im Rahmen dieser Arbeit zwei weitere Werkzeuge *SLAnalyzer* und *SLSmellsDetector* umgesetzt. *SLAnalyzer* bietet erweiterte Operationen zur Modellanalyse, während *SLSmellsDetector* strukturelle Schwachstellen (sogenannte „Model-Smells“) in einem Modell durchsucht. Allerdings werden diese Werkzeuge in dieser Veröffentlichung nicht mehr weiter vertieft.

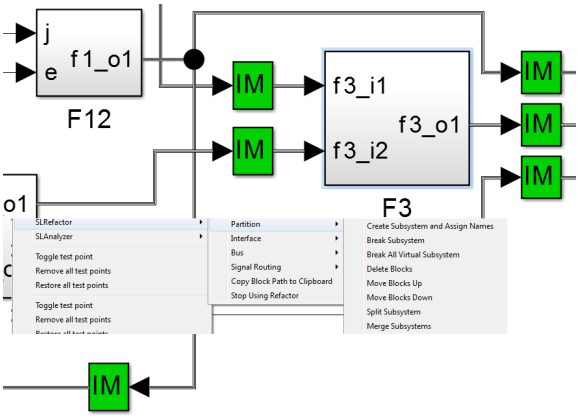


Abb. 5: Integration von SLRefactor im Simulink-Editor

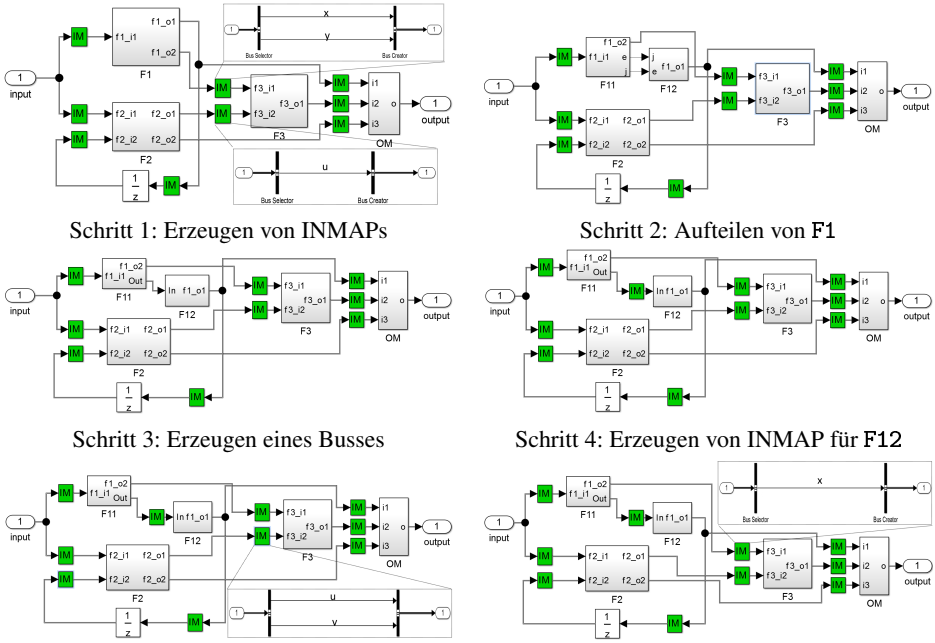
6 Beispiel: Umbau des Modells mit Hilfe von SLRefactor

In diesem Abschnitt wird durch ein Beispiel gezeigt, wie der Entwickler SLRefactor benutzen kann, um die Struktur des Modells sukzessiv effizient zu verbessern und zu erweitern. Das Hauptmerkmal des hier verwendeten Modellierungsmusters ist die Nutzung von INMAP-Subsystemen zur Darstellung von effektiven Schnittstellen der Module. Die im Folgenden beschriebenen sechs Modifikationsschritte werden zudem durch die Darstellungen in Abb. 6 unterstützt.

Schritt 1: Erzeugen INMAPs für die Module Zur Verdeutlichung der Signalflüsse werden im ersten Schritt für jedes Modul sogenannte *INMAP-Subsysteme* erzeugt, die alle im jeweiligen Modul verwendeten Signale explizit auswählen. Hierfür wird die Refactoring-Operation *Create Effective Input Interface* benutzt. Durch die Verwendung von INMAP-Subsystemen wird die sogenannte *effektive Schnittstelle* der Module sichtbar. Das bedeutet, dass die im Modul verwendeten Signale durch INMAP-Subsysteme explizit spezifiziert werden. Somit ist beispielsweise erkennbar, dass das Modul `F3` das Signal `u` von Modul `F2` verwendet. In einem realen Modell können hunderte Signale in einem INMAP-Subsystem selektiert und zu einem Bus zusammengefasst werden.

Schritt 2: Aufteilen von `F1` in zwei kleinere Module Das Modul `F1` implementiert zwei Funktionalitäten, die nach Einschätzung eines Entwicklers in jeweils eigene Module gehören. Mit Hilfe der Refactoring-Operation *Split Subsystem* kann `F1` in zwei kleinere Module `F11` und `F12` aufgeteilt werden. Dabei werden die Signalverbindungen der betroffenen Blöcke beibehalten. Mit Hilfe der Refactoring-Operation *Split Subsystem* kann `F1` in zwei kleinere Module `F11` und `F12` aufgeteilt werden. Dabei werden die Signalverbindungen der betroffenen Blöcke beibehalten.

Schritt 3: Erzeugen eines Busses Nach der Aufteilung von `F1` kommunizieren `F11` und `F12` über elementare Signale. Gemäß Modellierungsmuster sollen Module auf oberste Ebene allerdings über Busse kommunizieren. Aus diesem Grund werden die elementaren



Schritt 1: Erzeugen von INMAPs

Schritt 2: Aufteilen von `F1`

Schritt 3: Erzeugen eines Busses

Schritt 4: Erzeugen von INMAP für `F12`

Schritt 5: Hinzufügen eines Signalpfades

Schritt 6: Entfernen eines Signalpfades

Abb. 6: Schrittweiser Umbau des Modells

Signale zwischen `F11` und `F12` durch die Refactoring-Operation *Create Bus* zu einem Bus zusammengefasst. Die dabei entstandenen *BusCreator*- und *BusSelector*-Blöcke werden mittels der Refactoring-Operation *Move Blocks Down* in `F11` bzw. `F12` verschoben.

Schritt 4: Erzeugen von INMAP für F12 Das Modul `F12` hat einen Eingangsbus, aber kein INMAP-Subsystem, in dem die effektive Schnittstelle beschrieben wird. Mit Hilfe von *Create Effective Input Interface* wird ein INMAP-Subsystem für `F12` erstellt.

Schritt 5: Hinzufügen eines Signals Nun wird festgestellt, dass ein Submodul innerhalb von `F3` das Signal `v` von `F2` benötigt. Die Refactoring-Operation *Add Deep Bus Signal* führt das Signal `v` von `F2` zu dem Zielmodul. Dabei wird das Signal durch die INMAP-Subsysteme geleitet. In der Abbildung lässt sich erkennen, dass das Signal `v` in einem INMAP-Subsystem von `F3` selektiert und zu dem Bus hinzugefügt wird.

Schritt 6: Entfernen eines Signals Der Entwickler stellt fest, dass das Signal `y` von `F1` in `F3` nicht mehr benutzt wird. Mit Hilfe der Refactoring-Operation *Remove Deep Signal Backwards* kann `y` entfernt werden. Dabei hat der Entwickler die Möglichkeit zu entscheiden, bis zu welcher Stelle `y` entfernt wird. In der Abbildung wird gezeigt, dass das Signal `y` in einem INMAP-Subsystem von `F3` entfernt wird.

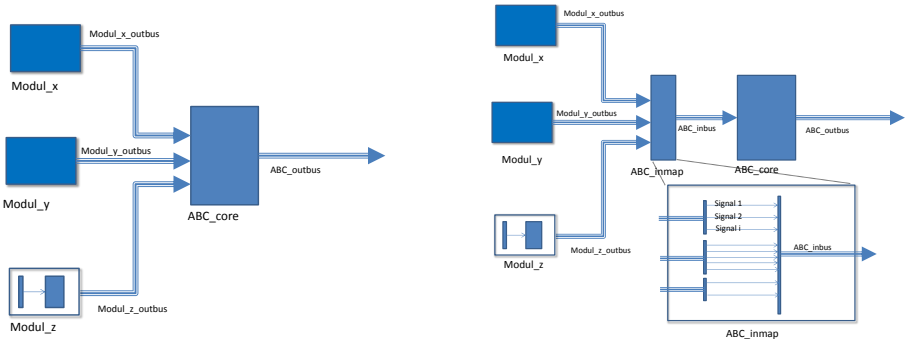
7 Verwandte Arbeiten

Eine Arbeit, die sich ausgiebig mit Transformation von Simulink-Modellen beschäftigt, ist das MATE-Projekt [Le10]. In diesem Projekt wurde untersucht, ob Graphtransformationstechniken für visuelle Spezifikation und Transformation von Simulink- und Stateflow-Modellen geeignet sind. In Vergleich zu der vorliegenden Arbeit liegt der Fokus des MATE-Projekts jedoch nicht auf der systematischen Erfassung von Refactoring-Operationen. Außerdem hat sich gezeigt, dass eine rein visuelle und auf Graphtransformationen basierende Spezifikationssprache für komplexe Spezifikationsszenarien nicht ausreichend ist. Beispielsweise ist die Spezifikation regulärer Ausdrücke, komplexer mathematischer Berechnungen sowie komplexer Navigation durch eine Netzliste verbundener Objekte nicht oder nur schwer möglich.

T. Gerlitz et al. [GSK13] schlagen ein Verfahren zur Duplikatserkennung auf Basis von Layoutinformationen vor. Darüber hinaus stellen sie ein Rahmenwerk vor, welches dieses Verfahren und verschiedene andere Duplikatserkennungsverfahren kombiniert, um höhere Präzision bei der Duplikatserkennung zu erreichen. Die gefundenen Duplikate können mithilfe einer Refactoring-Operation in einen gemeinsamen generischen Bibliotheksblock extrahiert werden. Die Duplikate werden durch einen entsprechend parametrisierten Bibliotheksblock ersetzt werden.

8 Feldstudie

Der hier vorgestellte Ansatz wurde über ca. zwei Jahre in verschiedenen Bereichen der Vor- und Serienentwicklung bei der Daimler AG erprobt. In regelmäßigen Gesprächen berichten die Entwickler, dass sie durch den Einsatz von SLRefactor Simulink-Modelle effizienter bearbeiten können. Im Folgenden wird anhand eines realen Anwendungsfalls die Effizienzsteigerung durch die Anwendung des Ansatzes demonstriert.



(a) Ausgangs-Modularchitektur (b) Modularchitektur nach Refactoring

Abb. 7: Modularchitektur des Modells

In Abb. 7a wird ein Ausschnitt der Modularchitektur des Ausgangsmodells des Fallbeispiels veranschaulicht. Der Entwickler erhält das Modell von einer Nachbarabteilung und möchte das Modell überarbeiten und erweitern. Hier werden Blöcke durch Module in Form

von Subsystemen funktional voneinander getrennt. Diese Module kommunizieren über Busse miteinander. In diesem Modell führen drei Busse in das Modul `ABC_core` hinein, die jeweils Signale der Module `Modul_x`, `Modul_y` und `Modul_z` übertragen.

Bei der Durchsicht des Modells stellt der Entwickler aber fest, dass das Modell strukturell nicht optimal ist. Insbesondere ist nicht sofort erkennbar, welche Signale der Busse tatsächlich im Modul `ABC_core` verwendet werden. Zur Verbesserung der Modellstruktur entscheidet der Entwickler, zunächst die potentiellen Schnittstellen auf *effektive Schnittstellen* mit Hilfe der Refactoring-Operation *Create Effective Input Interface* zu reduzieren, die nur die tatsächlich verwendeten Signale darstellen. Abb. 7b zeigt die Modularchitektur nach der Anwendung von *Create Effective Input Interface*. Außerdem möchte der Entwickler für jedes Modul ein einziges INMAP-Subsystem erzeugen. Damit werden die Eingangssignale des Moduls `ABC_core` so zusammengefasst, dass das Modul genau einen Eingangsbus mit allen in dem Modul verwendeten Signalen erhält.

In dem Fallspiel gibt es insgesamt 9 Module, die wie `ABC_core` Signale anderer Module über eingehende Busse verwenden. Tabelle 1a zeigt sowohl die Anzahl der Blöcke, als auch die Hierarchietiefe dieser 9 Module. Um das Modellierungsmuster mit INMAP

Modul	Σ Blöcke	Hierarchietiefe
1	1349	5
2	1046	4
3	697	4
4	233	3
5	1304	4
6	29	1
7	243	2
8	297	2
9	255	3

(a) Umfang der Module

Modul	Σ verwendete Signale
1	156
2	179
3	157
4	130
5	50
6	76
7	11
8	130
9	60
	949

(b) Umfang der INMAPs

Tab. 1: Statistische Daten

umzusetzen, müssen 9 INMAP-Subsysteme manuell erzeugt werden. Dabei müssen insgesamt fast 1000 Signale per Hand identifiziert werden, die in den jeweiligen Modulen verwendet werden (Tabelle 1b). Nach einer Expertenabschätzung benötigt ein Entwickler zur Durchführung der Änderung mehrere Stunden. Zudem ist die manuelle Änderung fehleranfällig. SLRefactor ermittelt die von den Modulen benötigten Signale automatisch und führt die Änderungen innerhalb weniger Minuten fehlerfrei durch. Danach kann der Entwickler durch die von SLRefactor bereitgestellten Refactoring-Operationen wie *Move Blocks Up/Down Split Subsystem*, *Merge Subsystems*, *Add Deep Signal*, *Remove Signal Backwards/Forwards* usw. die Struktur des Modells effizient verändern und erweitern.

9 Zusammenfassung und Ausblick

Im Rahmen der vorliegenden Arbeit ist erstmalig ein Refactoring-Konzept für Simulink-Modelle entwickelt worden. Das Konzept ist an bekannten Refactoring-Techniken für textuelle Programmiersprachen angelehnt. Hierzu wurde ein Katalog von Refactoring-Operationen erarbeitet. Das Konzept wurde in Form eines Prototyps umgesetzt, welcher in den Simulink-Editor integriert ist und die meisten Refactoring-Operationen aus dem Katalog bereitstellt. Durch eine Erprobung bei der Daimler AG konnte die Relevanz und Praxistauglichkeit des Ansatzes gezeigt werden. Der Ansatz beschränkt sich derzeit auf Simulink-Modelle. Grundsätzlich ist aber davon auszugehen, dass die Ideen auch auf andere vergleichbare datenflussorientierte grafische Modellierungssprachen übertragbar sind.

Literaturverzeichnis

[DRW12] Dziobek, Christian; Ringler, Thomas; Wohlgemuth, Florian: Herausforderungen bei der modellbasierten Entwicklung verteilter Fahrzeugfunktionen in einer verteilten Entwicklungsorganisation. In: Proceedings Dagstuhl-Workshop MBEES Modellbasierte Entwicklung eingebetteter Systeme. S. 1–10, 2012.

[Fo99] Fowler, Martin: Refactoring: Improving the Design of Existing Code. Addison-Wesley, 1999.

[Ga94] Gamma, Erich; Helm, Richard; Johnson, Ralph; Vlissides, John: Design patterns: elements of reusable object-oriented software. Pearson Education, 1994.

[GSK13] Gerlitz, Thomas; Schake, Stefan; Kowalewski, Stefan: Ansatz zur Erstellung und Wartung von Simulink-Modellen durch den Einsatz von Transformationen/Refactorings und Generierungsoperationen. In: Proceedings Dagstuhl-Workshop MBEES Modellbasierte Entwicklung eingebetteter Systeme. S. 1–10, 2013.

[Kl12] Klauske, Lars K.: Effizientes Bearbeiten von Simulink Modellen mit Hilfe eines spezifisch angepassten Layoutalgorithmus. Dissertation, Technical Universität Berlin, 2012.

[Le10] Legros, Elodie; Schäfer, Wilhelm; Schürr, Andy; Stürmer, Ingo: MATE: A Model Analysis and Transformation Environment for MATLAB Simulink. In: Proceedings of the 2007 International Dagstuhl Conference on Model-based Engineering of Embedded Real-time Systems. MBEERTS'07. Springer-Verlag, S. 323–328, 2010.

[Ma14] Mathworks, The: , Matlab/Simulink. URL `http://www.mathworks.com/toolbox/simulink/`, 2014.

[QD13] Quang, Tran M.; Dziobek, Christian: Ansatz zur Erstellung und Wartung von Simulink-Modellen durch den Einsatz von Transformationen/Refactorings und Generierungsoperationen. In: Proceedings Dagstuhl-Workshop MBEES Modellbasierte Entwicklung eingebetteter Systeme. S. 1–10, 2013.

[Ra02] Rau, Andreas: On model-based development: A pattern for strong interfaces in SIMULINK. Gesellschaft für Informatik, FG, 2(1):12, 2002.

[TWD13] Tran, Quang M.; Wilmes, Benjamin; Dziobek, Christian: Refactoring of Simulink Diagrams via Composition of Transformation Steps. In: Proceedings of the 8th International Conference on Software Engineering Advances (ICSEA). 2013.

[Wi06] Widmer, Tobias: Unleashing the Power of Refactoring. In: Eclipse Magazine. Software & Support Verlag, 2006.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Generating Multi-objective Programs from Variant-rich EAST-ADL Product Line Architectures

Tobias Wägemann[1] and Albert Werner[2]

Abstract: The design of robust architectures for software-intensive systems in the automotive domain is a complex task and a considerable engineering challenge. Even conventional systems are usually subject to a multitude of conflicting design goals, such as unit cost and weight minimization, dependability augmentation and timing control. One feasible approach to resolve these conflicting levers on a system's architecture is to perform an optimization analysis on a concretely defined design space, which in general is vast. For the purpose of the approach presented here, this design space is represented by an automated identification of variation points relevant for design space exploration. The identified variability information is then transformed into a convenient mathematical representation for product-line-aware architecture optimization.

Keywords: Model-based Analysis, Multi-objective Programming, Architecture Optimization, Pareto Optimality, Product Line Engineering

1 Introduction

The field of automotive engineering has seen a steady increase in relevance of software-intensive electronic systems over the last decade. In the early stages of this development, software-based components were mainly used for replacing or complementing electromechanical functionality. More modern approaches have instead striven to explore entirely new application areas within the scope of the automotive domain. As a consequence, the complexity of employed system architectures has grown rapidly, including those architectures of safety-critical applications and controls.

The current challenge in this field of research is therefore not merely the exploration of additional use cases for software applications, but also the optimization of existing and future system architectures taking a multitude of different quality attributes, i.e. design objectives, into consideration [Wa13]. This challenge is further complicated by the fact that many real-world objectives conflict with each other, which necessitates a mechanism for trade-off analysis and resolution.

The resulting trade-off analysis is a hard combinatorial problem, which can often not be satisfactorily resolved by hand for all but the most trivial cases. In absence of adequate automated solutions, industry standard often relies solely on manual decision making when

[1] Technische Hochschule Nürnberg Georg Simon Ohm, Fakultät Informatik, Hohfederstr. 40, 90489 Nuremberg, Tobias.Waegemann@th-nuernberg.de
[2] Technische Hochschule Nürnberg Georg Simon Ohm, Fakultät Informatik, Hohfederstr. 40, 90489 Nuremberg, Werneral43493@th-nuernberg.de

it comes to finding optimal or near-optimal configurations for variant-rich architectures. Due to the complex interdependencies and decision ramifications in sufficiently large design spaces, it is reasonable to assume that a large number of currently employed system designs in the automotive domain are based on potentially suboptimal configuration decisions.

In order to establish a robust foundation for automated optimization approaches, a system's information model has to facilitate a high degree of expressiveness in regard to its variability management capabilities and its support of relevant quality attribute data. The utilization of a domain-specific ADL like EAST-ADL[3] allows for model-based analyses on an information model of ample expressiveness comprised in a single monolithic model structure [Bl13]. A comprehensively engineered EAST-ADL model can effectively provide all relevant information necessary for an optimization analysis; including the system's variability information and a bulk of data eligible for quality attribute assessment.

To make use of such information repositories for the purpose of architecture optimization, our goal is to devise an approach for deriving multi-objective programs from variant-rich EAST-ADL models. This can be realized by representing quality attribute data as objective functions and variability information as program constraints. We are also taking care in creating a heuristic variability interpretation process, which allows for product-line-aware optimization analyses, i.e. the possibility of family-based optimization of entire product lines in one sweep, as opposed to the optimization of individual products [Th12].

In this paper, we provide a summary of our ongoing work on the subject of generating multi-objective programs from variability information and quality attribute data of given EAST-ADL models and detail our progress in regard to different aspects of the intended implementation.

2 Variant-rich EAST-ADL System Architectures

The EAST-ADL is an architecture description language with an explicit focus on the specific requirements of architecture design and management in the automotive domain. The language was developed and refined in a number of international research projects, notably ITEA EAST-EEA (www.itea3.org/project/east-eea.html), ATESST and ATESST2 (www.atesst.org), and MAENAD (www.maenad.eu), and is currently being overseen by the EAST-ADL Association (www.east-adl.info). The language has been tailored towards compatibility with the well-established AUTOSAR standard (www.autosar.org), which in turn serves as an integral part of the EAST-ADL language by realizing one of its abstraction layers outright. The purpose of the EAST-ADL is to provide a standardized solution for creating exhaustive representations of software-intensive electronic systems in the automotive domain, by introducing additional layers of abstraction above the software architecture centric perspective of AUTOSAR. EAST-ADL models can further be enriched by language-inherent constraints for requirements and other supplementary information.

[3] *Electronics Architecture and Software Technology - Architecture Description Language*, an ADL specifically tailored to the requirements and characteristics of the automotive domain.

2.1 The EAST-ADL System Model

The EAST-ADL system model is segmented into four vertical abstraction layers, which can be augmented horizontally by an environment model and several language extensions (cf. Figure 1). Each layer has a specific purpose in the overall language concept and represents a different technical abstraction viewpoint, as well as a separation of concerns during the development process [EA13]. The EAST-ADL abstraction levels are as follows:

The *Vehicle Level* represents a top-level view of the system's characteristics and functions, as defined by stakeholder requirements, by means of cardinality-based feature modelling [CK05]. The focal point of this abstraction level is the technical feature model, which has a markedly technical perspective on the system's functional and non-functional composition. This technical feature model can be complemented by additional feature models, which realize different views on the composition of the system or any of its subsystems. The scope of these additional feature models can correspond to the requirements and practices of different departments or divisions, e.g. a specialized view for marketing purposes.

The *Analysis Level* implements an abstract functional description of the system as a functional analysis architecture (FAA), without exposing any implementation details. The FAA is comprised of analysis functions, which in conjunction represent the system's functional behavior in accordance with the specifications defined in the technical feature model. This kind of high-level functional decomposition allows for verification and validation (V&V) tasks on a purely functional model, without interfering effects caused by implementation details.

The *Design Level* is divided into two parts: a functional design architecture (FDA) and a hardware design architecture (HDA). The FDA defines a concrete functional decomposition of the system's application software structure, with the intention of meeting certain non-functional constraints such as function-to-hardware allocation and reuse considerations. The FDA therefore comprises concrete design functions and and local device managers in a way that facilitates a derivation of AUTOSAR software components (SWCs), potentially using an n:m mapping. The HDA defines the system's hardware design by means of logical hardware functions, in order to establish a concrete representation of the physical device layout and circuitry, and to allow for allocation of design functions to modelled hardware entities.

The *Implementation Level* does not have an innate EAST-ADL implementation, instead the AUTOSAR standard is used for creating a suitable software architecture representation for the system. It should be noted that the EAST-ADL supports vertical traceability throughout all abstraction layers, including the Implementation Level realized by AUTOSAR.

The language extension components can be used to represent supplementary information, including requirements, variability information, timing and dependability constraints and generic system constraints, across all layers of an EAST-ADL architecture. An environment model, also called plant model, can be used to represent elements which influence and interact with the system architecture, but are out of scope of the system model itself, e.g. the near environment of the vehicle including other nearby vehicles.

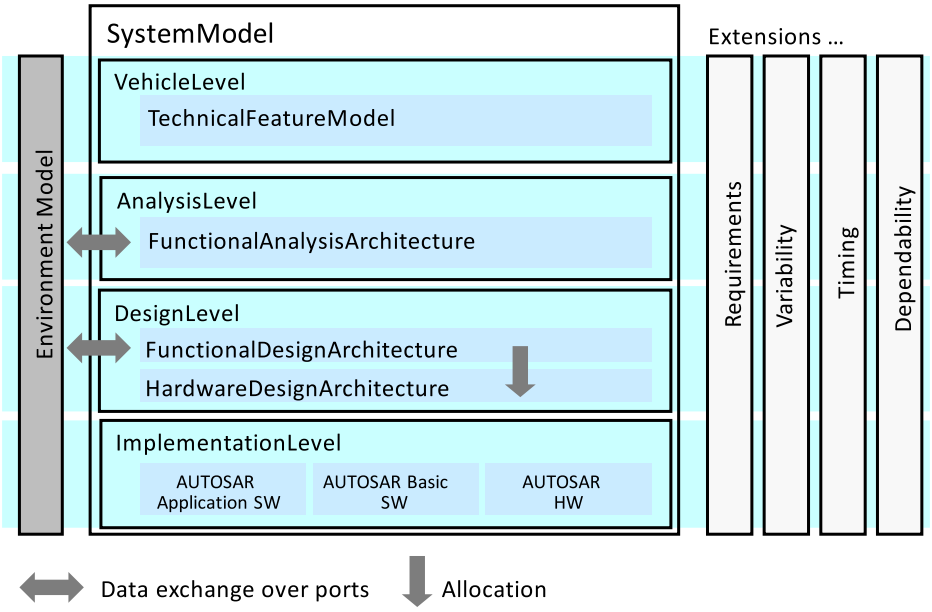


Fig. 1: The EAST-ADL System Model [Bl13, p.4]

2.2 Variability Language Concepts

Variability modelling in EAST-ADL can generally be divided into two different categories. The one category, variability on the Vehicle Level, is represented in the form of *cardinality-based feature models* (CBFM) [CK05]. Its purpose is to establish an abstract view of the system's overall variable content, including the interdependencies between variable elements. This view offers a distinction between a technical perspective and customer-specific perspectives, but does not disclose any particulars about actual implementation details. The other category, variability on the artifact levels, can be represented as part of the systems FAA, FDA, and HDA models by means of both CBFM and concrete variation points. This view is concerned with the actual technical implementation of variable content, by defining the relationships and interdependencies between variable architecture entities, i.e. actual functional components on the artifact levels of the system architecture.

In order to relate variability information of different feature models, as well as to allow for logical bindings from feature models to variation points on the artifact level, the language also supports configuration modelling by means of configuration links. These links contain atomized rules – so called configuration decisions – for the configuration of target model entities based on a given configuration of source model entities. Configuration links can be used to relay configuration information across EAST-ADL layers; but also in a more local context, for creating bindings of the internal variability of a container element with feature model representations of that internal variability – a so called *public feature*

model. The application of these variability concepts facilitates the use of *compositional variability management* (CVM) as a systematic language concept of the EAST-ADL language [RKW09, Re09].

2.3 Product Line Variability versus Architecture Design Space

In addition to the observance of different variability description methods described in section 2.2, a distinction between product-line-oriented variability [MP07] and the system's architecture design space, also called *architectural degrees of freedom* [Al13], is essential for realizing a product-line-aware architecture optimization approach. Table 1 gives a summary of the conceptual differences between these two kinds of variability.

Product Line Variability	Architecture Design Space
Purpose: satisfies specific needs of a particular market segment or mission (e.g different models of cars).	Purpose: represents different implementation variants (e.g. multiple suppliers for a subsystem).
Variations have functional influence on the target system.	Variations have influence on system characteristics like cost, performance or dependability.

Tab. 1: Characteristics of product line variability and architecture design space

The intended output of our proposed optimization approach is a product line with optimal architectural decisions, as opposed to an optimally configured product based on the product line. It is therefore necessary to establish a way to differentiate between the two types of variability, in order to be able to mark product-line-oriented variations as restricted for optimization. The approach presented in this paper achieves this distinction by tracing the propagation of individual variation points through the different layers of the given model.

In case of product line variability, a regarded variation point can always be traced all the way up to the features on the model's Vehicle Level (cf. section refsubsec:eadlSysModel), by means of configuration decision connections. The features represent product line variations and therefore allow for configuration of variation points located on underlying abstraction levels, i.e. all variability that ultimately is rooted on Vehicle Level is product line variability.

If no such traceable connectivity to the Vehicle Level exists, the variation point is instead considered to be part of the model's architecture design space. Such variation points are not configured as part of a product line configuration, but are instead decided at the system design time. Finding optimal allocations for these architecture variation point is the primary objective of the intended optimization approach.

3 Architecture Optimization Approach

In order to establish a robust foundation for optimization analyses of variant-rich EAST-ADL models, an expedient course of action is to first educe a rigorous mathematical model,

which reduces the complexity of the initial problem down to the fundamental drivers of the optimization process. This distillation procedure results in a lean representation of the optimization problem, which can be easily transformed into fitting input data for existing optimization applications.

In EAST-ADL models, however, the information relating to variability and design objectives is interwoven with other aspects of the given architecture description. There exists no separate depository for this kind of supplementary data; the identification and interpretation of the relevant information together with the transformation into a proper mathematical formulation is the purpose of the research project presented in this paper.

In order to accomplish this goal, we propose an approach for generating multi-objective programs from given EAST-ADL models, by representing the design objectives as a set of objective functions, and the system design space as a set of program constraints. Figure 2 illustrates the basic workflow of deriving such a program from a given model in terms of relevant activities. In order to allow for product-line-aware optimization, a number of specific considerations have to be made in regard to the examined variability information (cf. section 2.3).

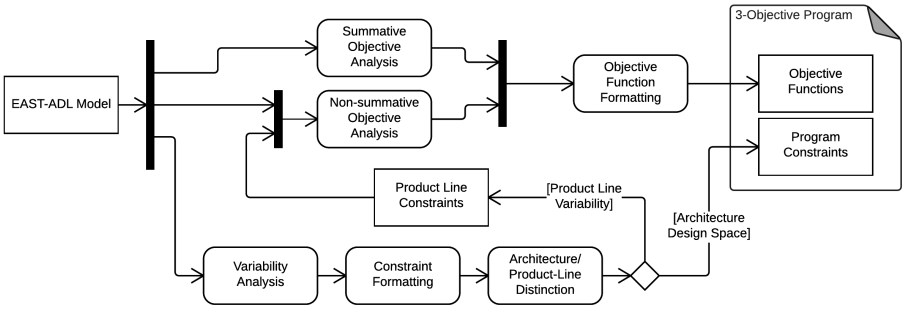


Fig. 2: Deriving a multi-objective program with both summative and non-summative design objectives

Our efforts to devise a comprehensive approach are part of an ongoing work in progress; our current, preliminary method still has certain restrictions in terms of the kind of variability and the kind of design objectives that can be handled and successfully converted. These restrictions are (a) the limitation to strictly boolean variability spaces, as opposed to non-boolean aspects introduced in particular by parametrization and cloned features, and (b) the limitation to strictly summative design objectives, like unit cost and weight minimization. The effect of these limitations on the generated output is that the resulting program is both linear and disjunctive. Future iterations, assuming that the current limitations can be lifted, may instead produce multi-objective non-linear mixed-integer programs (cf. section 5).

3.1 Multi-objective Optimization

When considering multiple objectives as part of an optimization analysis, it is usually impossible to find a single perfect solution which is truly optimal with respect to all

objectives. This complication is caused by the often conflicting nature of typical design objectives, e.g. improving safety will decrease performance, and vice versa. One possible solution for this dilemma is to convert the multi-objective analysis into a single-objective analysis, by formulating an entirely new objective function from a weighted aggregation of the original objectives – a process known as scalarization or weighted normalization [GR06]. This approach, while widely used, has significant drawbacks, in particular the fact that the weighting factors for the objectives have to be defined beforehand, i.e. before there is tangible knowledge about the properties of possible optimal solutions.

An alternative and more exhaustive approach involves the simultaneous evaluation of all distinct objective functions in a single optimization sweep. This approach is usually regarded as genuine multi-objective optimization, also called *Pareto* optimization due to the particular composition of the analysis result. The typical outcome of a Pareto optimization is a Pareto set, also called Pareto frontier, which constitutes the set of all Pareto-optimal solutions of the optimization analysis [MA12b, p.54f].

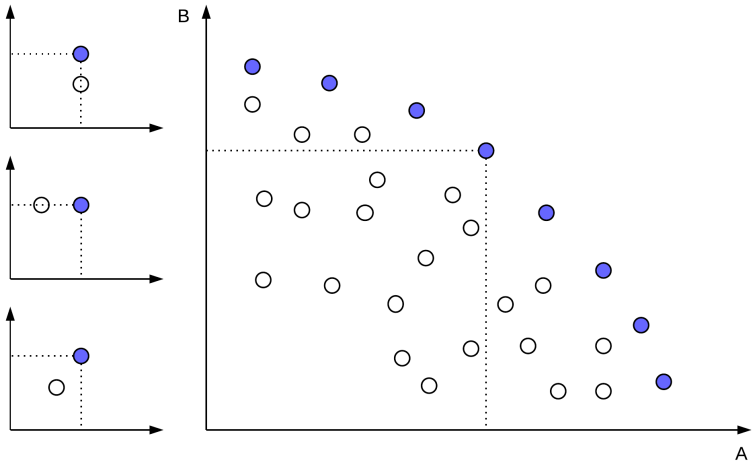


Fig. 3: Depiction of Pareto dominance for an exemplary solution space

Pareto optimality is based on the concept of dominance. A solution is called dominant if and only if no other solutions exist, which outperform it with respect to at least one objective without degradation in any other objective [BK05, p.414ff]. Figure 3 depicts a solution space for an optimization analysis with two conflicting design objectives A and B plotted against the axes. In this example, both objective functions were to be maximized – think safety and performance. The Pareto set is represented by the filled bullets in the main graph. The smaller graphs depict different scenarios of solution dominance.

3.2 Architecture Representation as Multi-objective Program

In order to attain a suitable mathematical representation of a given EAST-ADL optimization task, we propose an approach which makes use of the formulations of multi-objective programming [Be71][MA04]. Our current exclusion of non-boolean variabilities and non-

summative design objectives, as noted in section 2, results in a program in which the constraints and objective functions are (a) linear, and (b) strictly binary. The application of such programs is called 0-1 integer linear programming, or *disjunctive programming* [Ba79]. Such a program can be formulated as follows:

$$\begin{aligned} \text{Minimize} \quad & Cx \\ \text{subject to} \quad & Ax \geq a_0 \\ & x \in \{0,1\} \end{aligned} \tag{1}$$

where C is a (m,n)-Matrix representing a mapping of objective coefficients; A and $a0$ are a (p,n)-Matrix and a p-vector, respectively, representing the mapping of the variability space to constraint coefficients; and x is an n-vector of strictly 0-1 binary architecture decisions.

The Matrix Cx is in principle just an alternative representation of a set of linear functions $F(x) = (f_1(x), f_2(x), ..., f_m(x))^T$, where each function represents one distinct design objective, and thus one of the target parameters in the Pareto trade-off analysis. The coefficients can be parsed directly from extension constraints affiliated to the respective variable architecture elements in the given EAST-ADL model (cf. section 2.2). The methodology of ascertaining the program constraints $Ax \geq a0$ from the model's architecture design space is highlighted in section 3.3.

In order to illustrate the composition of such a program, suppose an exemplary optimization task on a system with two variable elements X_1 and X_2, which are mandatory alternatives, i.e. exactly one of the elements can and must exist in each valid system configuration. Further suppose that the relevant design objectives for this optimization task be unit cost and weight, and that the elements have the properties $X_1(weight : 150, cost : 20)$ and $X_2(weight : 100, cost : 30)$. This supposition results in the following 2-objective program:

$$\begin{aligned} \text{Minimize} \quad & 150x_1 + 100x_2 \\ \text{Minimize} \quad & 20x_1 + 30x_2 \\ \text{s.t.} \quad & x_1 + x_2 = 1 \\ & x_1, x_2 \in \{0,1\} \end{aligned} \tag{2}$$

where each of the objective functions aims to minimize the regarded design objective unit cost and weight, respectively, and the program constraints ensure that one and only one of the elements exists once and only once in every solution.

Each of the Pareto-optimal solutions to this mathematical program represents a corresponding Pareto-optimal configuration of the initial system. It should be noted that, while the scale of constraints is rather handy in this small example, the size and complexity of program constraints grows rapidly with increasing complexity of a regarded system's variability space and might be problematic when considering very large systems. This is one of the so far unresolved challenges for future work on this project (cf. section 5).

3.3 Generation of Program Constraints from System Design Space

The basis for our optimization space definition is a variant-rich EAST-ADL model in the standardized exchange format EAXML, which is specified by the EAST-ADL Association by means of an XML schema[4]. The hierarchical XML structure of the input model can be transformed into a DOM tree in order to apply recursive search and pattern recognition operations. The format of the generated optimization constraints is conform to a set of program constraints for a disjunctive program. Taken together, these constraints are a representation of all valid configurations of the optimization space.

Constraints regarding features connected by configuration decisions follow the notation of the Variability Specification Language (VSL) [Re09]. Unlike artifacts, feature names are not identified by their fully qualified names but with a combination of the feature model's name and the feature's name separated by the hash symbol '#'. If the feature has a parent feature, the resulting path is arranged as follows: Featuremodel#Parentfeature.Childfeature (cf. Figure 4). Additionally, the VSL language allows for several alternative notations[5]. To get a consistent format, all VSL-based feature names are replaced by their fully qualified names and notations are replaced by a uniform version.

Next, the fully qualified names of variation points are substituted with simplified versions. These substitutions have the advantage of containing no special characters (e.g. '/'), which might otherwise cause problems during execution of the optimization procedure. In order to be able to re-obtain already placed substitutions, an assignment table is generated as part of this procedure. This step is necessary to avoid creating duplicates with differing name substitutions.

After these preparations have been finished, the variability information can be parsed into program constraints. Table 2 gives a summary of transformation rules for different kinds of variability modelling concepts. Applying these rules on the example shown in Figure 4, the variability information is first transformed into propositional logic. The example showcases a decision point regarding two alternative realizations of a breaking system, where one of the variants is dependent on the inclusion of the ABS subsystem. In the appendant feature model, all features are stated with their fully qualified names, which are later replaced by abbreviations.

Propositional logic for single features:

BS			Braking System is always present
b	$\rightarrow$	BS	when basic, then also Braking System
a	$\rightarrow$	BS	when advanced then also Braking System

Propositional logic for the feature group:

BS	$\rightarrow$	$(a\|b)$	when Braking System, then advanced or basic or both
BS	$\rightarrow$	$(!b\|!a)$	when Braking System, then not advanced and basic together

Propositional logic for the feature link:

a	$\rightarrow$	ABS	when advanced then ABS

[4] http://www.east-adl.info/Specification.html

[5] Different variations of operators (e.g. and, &, $\wedge$)

Next, the propositional logic is transformed into program constraints according to the second set of rules located in the third column of Table 2:

$$\begin{array}{lcl} BS & = & 1 \\ BS-b & \geq & 0 \\ BS-a & \geq & 0 \\ a+b-BS & \geq & 0 \\ -a-b-BS & \geq & -2 \\ ABS-a & \geq & 0 \end{array}$$

The set of these constraints, together with the elements' co-domain $BS,a,b,ABS \in \{0,1\}$, constitutes the set of all valid system configurations for this exemplary system.

Variability Method		Propositional Logic	Disjunctive Integer Program
Feature Tree	Feature has parent	$f \rightarrow f_{parent}$	$f_{parent} - f \geq 0$
	Feature is mandatory	f	$f = 1$
	Feature is excluded	$!f$	$(1-f) = 1$
	Feature Group	$\bigwedge_m \left(f_{parent} \rightarrow M_m(f_1,..,f_n)\right)$,	$\bigwedge_m \left(M_m(f_1,..,f_n) - f_{parent} \geq 0\right)$[6]
Feature Link	needs	$f_{start} \rightarrow f_{end}$	$f_{end} - f_{start} \geq 0$
	optional alternative	$f_{start} \rightarrow !f_{end}$	$f_{end} + f_{start} \leq 1$
	mandatory alternative	$f_{start} \oplus f_{end}$	$f_{start} + f_{end} = 1$
Variation Group	needs	$f_1 \rightarrow (f_2 \wedge f_3 \wedge \ldots \wedge f_n)$	$\bigwedge_{k=2}^{n} (f_k - f_1 \geq 0)$
	optional alternative	$\bigwedge_m (M_m(f_1,..,f_n))$	$f_1 + f_2 + \ldots + f_n \leq 1$
	mandatory alternative	$f_1 \oplus f_2 \oplus \ldots \oplus f_n$	$f_1 + f_2 + \ldots + f_n = 1$
Configuration Decision		$criterion \rightarrow effect$	$effect - criterion \geq 0$

f	: feature or artifact	!	: boolean operand 'not'
f_{parent}	: parent feature	$\oplus$	: boolean operand 'xor'
$\rightarrow$	: boolean operand 'implication'	$\bigwedge_m$	: for all m
f_{start}	: start feature regarding a dependency	f_{end}	: end feature regarding a dependency
$\wedge$	: boolean operand 'and'	$\bigwedge_{k=2}^{n}$	: for all $k \in \{2,\ldots,n\}$
M_m	: Maxterms[7] of all illegal combinations of features f_n in a feature group		

Tab. 2: Transformation rules

[6] Negated variables are replaced with the transformation rule for excluded features.
[7] Maxterms are true only for one single combination of all variables.

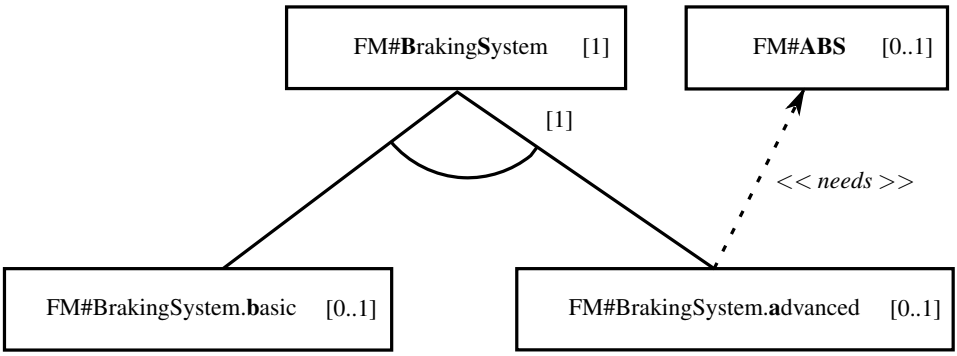


Fig. 4: Example feature model

4 Results and Discussion

The presented approach for deriving multi-objective programs from variant-rich EAST-ADL models has been shown to work for a particular subset of models with boolean variability spaces and summative design objectives. The developed mechanism is able to achieve this result by (a) formulating a set of *N* linear objective functions for *N* considered quality attributes, and by (b) formulating a set of linear program constraints, which together constitute a comprehensive representation of a given model's architecture design space.

By means of additional analyses on the model's variability space, the mechanism can further distinguish between product-line-oriented and architecture-oriented variation points, thereby allowing for product-line-aware optimization. In order to achieve the same capability for non-summative design objectives, we expect that the separated product line variability has to also be considered as part of the formulation of the – then potentially non-linear – objective functions (cf. section 5).

In its current iteration, the approach has proven effective when applied to the particular subset of EAST-ADL models as curtailed in this paper. Due to the flexibility of the utilized mathematical representation, the approach also shows promise in regard to future expansions for considering the full range of variability options in the EAST-ADL language, as well as a coverage of more complex design objectives – in particular non-summative objectives.

5 Outlook

At the time of release of this paper, the presented approach has only been realized for a particular subset of all feasible variant-rich EAST-ADL models, namely those with strictly boolean variability spaces and strictly summative design objectives. In order to expand the approach to cover a broader range of use cases, a number of changes have to be made to the existing mechanism.

One intended goal for future iterations is the support of non-boolean EAST-ADL-based variability information, in particular parametrization and cloned features. Since these

extensions can give rise to non-binary decision variables, it essentially entails the subsequent implementation of an additional change, namely the explicit evaluation of the respective co-domain of each decision variable.

Other projected extensions are concerned with realizing the interpretation and transformation of non-summative quality attribute data into corresponding objective functions, e.g. safety characteristics or performance constraints. We expect that most non-summative design objectives will have to be addressed by individually tailored approaches. Therefore, a reasonable first step for such an effort could be a real-world evaluation of the relative relevance of different quality attributes in the domain of automotive software-intensive systems.

Once a complete or near-complete coverage of variability characteristics and relevant design objectives has been established, a reasonable next step would be a conversion of the analysis output to a standardized optimization format like MPS. Existing tools for multi-objective programming could then be used to conduct case studies on the feasibility of our approach as a basis for optimization analyses on large test models. A sound implementation of this functional coverage would also be a useful contribution to the official EAST-ADL tool platform EATOP [MA12a].

Acknowledgments

We want to express our sincere gratitude and appreciation to Prof. Dr. Ramin Tavakoli Kolagari of the Technische Hochschule Nürnberg Georg Simon Ohm, and Prof. Dr. Klaus Schmid of the University of Hildesheim, for their continuous support in our research efforts and for sharing with us their knowledge and enthusiasm for variation management and automotive software engineering in general.

References

[Al13] Aleti, Aldeida; Buhnova, Barbora; Grunske, Lars; Koziolek, Anne; Meedeniya, Indika: Software architecture optimization methods: A systematic literature review. In: Software Engineering, IEEE Transactions on. volume 39. IEEE, pp. 658–683, 2013.

[Ba79] Balas, Egon: Disjunctive programming. In: Annals of Discrete Mathematics. volume 5. North-Holland Publishing Company, pp. 3–51, 1979.

[Be71] Benayoun, R; De Montgolfier, J; Tergny, Jo; Laritchev, O: Linear programming with multiple objective functions: Step method (STEM). In: Mathematical programming. volume 1. Springer, pp. 366–375, 1971.

[BK05] Burke, Edmund K; Kendall, Graham: Search methodologies: Introductory Tutorials in Optimization and Decision Support Techniques. Springer, 2005.

[Bl13] Blom, Hans; Lonn, Henrik; Hagl, Frank; Papadopoulos, Yiannis; Reiser, Mark-Oliver; Sjostedt, Carl-Johan; Chen, De-Jiu; Kolagari, Ramin Tavakoli: , White Paper Version 2.1.12: EAST-ADL - An Architecture Description Language for Automotive Software-Intensive Systems, 2013.

[CK05] Czarnecki, Krzysztof; Kim, Chang Hwan Peter: Cardinality-based feature modeling and constraints: A progress report. In: International Workshop on Software Factories. pp. 16–20, 2005.

[EA13] EAST-ADL Domain Model Specification, 2013. Version V2.1.12. http://east-adl.info/Specification/V2.1.12/EAST-ADL-Specification_V2.1.12.pdf.

[GR06] Grodzevich, Oleg; Romanko, Oleksandr: Normalization and other topics in multi-objective optimization. In: Proceedings of the Fields–MITACS Industrial Problems Workshop. The Fields Institute, pp. 89–102, 2006.

[MA04] Marler, R Timothy; Arora, Jasbir S: Survey of multi-objective optimization methods for engineering. In: Structural and multidisciplinary optimization. volume 26. Springer, pp. 369–395, 2004.

[MA12a] MAENAD consortium: , EATOP: An EAST-ADL Tool Platform for Eclipse, 2012. Project Deliverable D5.3.1 Version 3.0. http://www.maenad.eu/public/Deliverables/MAENAD_Deliverable_D5.3.1_EATOP_V3.0.pdf.

[MA12b] MAENAD consortium: , Language Concepts Supporting Engineering Scenarios, 2012. Project Deliverable D3.1.1 Version 3.0. http://www.maenad.eu/public/Deliverables/MAENAD_Deliverable_D3.1.1_V3.0.pdf.

[MP07] Metzger, Andreas; Pohl, Klaus: Variability management in software product line engineering. In: Companion to the proceedings of the 29th International Conference on Software Engineering. IEEE Computer Society, pp. 186–187, 2007.

[Re09] Reiser, Mark-Oliver: Core Concepts of the Compositional Variability Management Framework (CVM): A Practitioner's Guide. Technical report, Technische Universität Berlin – Fakultät IV – Softwaretechnik, 2009.

[RKW09] Reiser, Mark-Oliver; Kolagari, Ramin Tavakoli; Weber, Matthias: Compositional variability-concepts and patterns. In: System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on. IEEE, pp. 1–10, 2009.

[Th12] Thüm, Thomas; Apel, Sven; Kästner, Christian; Kuhlemann, Martin; Schaefer, Ina; Saake, Gunter: Analysis strategies for software product lines. Technical report, Otto-von-Guericke-Universität Magdeburg Fakultät Informatik, 2012.

[Wa13] Walker, Martin; Reiser, Mark-Oliver; Tucci-Piergiovanni, Sara; Papadopoulos, Yiannis; Lönn, Henrik; Mraidha, Chokri; Parker, David; Chen, DeJiu; Servat, David: Automatic optimisation of system architectures using EAST-ADL. Journal of Systems and Software, 86(10):2467–2487, 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Agile Synchronization between a Software Product Line and its Products

Bernhard Rumpe [1]Christoph Schulze [1] Johannes Richenhagen [2] Axel Schloßer [2]

Abstract:

Establishing and maintaining a software product line for a series of similar applications is a complex and time-intensive process, which can only pay back its cost if the provided software components achieve a high degree of reuseablity. Furthermore, in many domains, including automotive, stability will always be prioritized over reusability. Integrating a set of existing similar but individually developed products into a reusable product line is often very time-intensive. It is better, but still time-consuming, to feed the result of a new product back into an already existing software product line. Due to resource constraints for the next product, necessary resources are rarely provided. As a consequence, a product line is often endangered by becoming outdated and thus less usable. Therefore, the establishment and maintenance of a software product line needs to be supported by processes and techniques which allow to derive necessary information from running projects with only minor or ideally non-manual effort. In this paper, an agile and semi-automated process to extract and maintain a software product line during the parallel evolvement of several products is proposed. This process is based on PERSIST, an industrially used approach which combines agile techniques and sustainable, long-term architecture development. At the end of this paper, the current status and achievements of PERSIST are discussed, already implemented techniques are evaluated in the context of the automotive powertrain, and further potentials are highlighted.

1 Introduction and Motivation

In the automotive domain, the complexity of software functions and the demanded quality standards, e.g. CMMI, ISO 26262, are still growing, while in the meantime the expected release cycles get shorter and shorter [Br06]. In addition, the vehicle starts to turn into a smart device which is able to interact with its environment and to react autonomously. As a consequence, further aspects, such as security or privacy, receive a higher prioritization [Pr07]. Features, e.g. connectivity, which seem to be of no importance in the automotive domain, may become mandatory for the customer, while main selling points of the past, e.g. driveability, could switch to secondary priority if not connected to the digital world properly. In parallel the complexity of the general required functionality of the automotive powertrain is increased, as more and more standards have to be fulfilled.

[1] Software Engineering, RWTH Aachen, Ahornstraße 55, 52074 Aachen, http://www.se-rwth.de
[2] FEV GmbH, Neuenhofstraße 181, 52078 Aachen, http://www.fev.com

This up-to-date scenario is a great example for the frequently changing and hardly predictable demands in today's automotive industry. Who can ensure that features which are identified as necessary today will be required in a similar form tomorrow? What is the adequate time frame for long-term strategies in the context of shorter and shorter release periods? How can already established software structures be connected to new sub-domains and thereby keep a maximum degree of reusability?

In [RPS14], PERSIST (powertrain control architecture enabling reusable software development for intelligent system tailoring) has been established to give first answers to these questions in the context of an engineering service supplier. This approach introduces agile methods to be able to react flexibly on requirement changes and to reduce the duration of one development cycle and project quality risks due to continuous integration. In addition standards like AUTOSAR and the ISO 25010 have been taken into consideration during the development of PERSIST to ensure a high degree of quality and to be able to comply to requested formats in todays automotive industry. PERSIST focuses on small development cycles in which a small amount of software components are planned, realized and verified. These components are verifed automatically by the nightly build server mostly. The approach was successfully applied first in the context of a two-stage turbocharged gasoline engine [RPS14]. Based on this experience, further applications have been performed [VRP15, Ri15].

Nevertheless, the establishment and maintenance of a Software Product Line (SPL) in the context of the daily work of a supplier has to face several ongoing projects in parallel. Each of these projects have similar demands, but different customers. Often these customers are only interested in the development of a product rather than in the establishment of reusable software components or a complete software product line, at least not in the degree it would be sufficient for the supplier.

As a consequence, an approach is required which allows the project teams to focus on the implementation of the required product, while it is possible to continuously establish and maintain a SPL in parallel with minimized effort. Therefore, it is necessary to further extend the collaboration between Agile Software Development (ASD) and Software Product Line Engineering (SPLE), resulting in Agile Software Product Line Engineering (APLE).

In this paper, we propose such an approach based on the already established fundament of PERSIST. The paper is structured as follows: in Section 2 we give a short overview about SPLE and ASD to be able to give a more detailed overview of the current status in the research field of APLE. The section finishes with summarizing metrics which build the foundation for the proposed approach. Section 3, describes the proposed approach in all detail, while in Section 4, its application is evaluated. Section 5 summarizes related work and in Section 6 open points are discussed and future work is derived.

2 Foundations

SPLE [PBL05] focuses on establishing a reusable platform for a specific domain which allows to derive several customized products in an efficient manner (Figure 1). The SPLE

paradigm divides the development into two separate phases: During the Domain Engineering (DE) phase the most common aspects of the domain are identified and based on this analysis a software platform is derived. During this step, the commonality and variability of the product line are defined. One important step is the definition of a reference architecture which represents the common high-level structure of the product line.

Based on this established platform, domain artifacts are reused and open variation points are bound to extract a specific product during the Application Engineering (AE) phase [PBL05]. To express variability in domain artifacts different annotative, compositional, and transformational variability modeling approaches have been proposed [Sc12].

A SPL can be established either in a proactive, reactive or extractive manner [Kr02]. The proactive approach derives the whole product line from scratch, while the reactive approach derives the product line in an incremental manner. In many cases, the initial product line is derived from a given set of products, therefore an extraction is performed.

ASD unites a set of methods which are build upon the Agile Manifesto [1]: *priorizing individuals and interactions, working software, customer collaboration and responding to change over processes and tools, comprehensive documentation, contract negotiation and following a plan*. This is achieved by performing small development cycles which are driven by related acceptance criteria. Long-term analysis and development plans are avoided as frequent changes are expected. SCRUM [Sc04] or extreme Programming [BA04] are popular software development processes in the context of ASD, while Test Driven Development (TDD) [Be02] and continuous integration are two of the most common agile methods [GPM08].

SPLE and ASD both pursue at the same targets: customer satisfaction and reduced time-to-market [Dí11]. It has already been demonstrated that both approaches are able to achieve the specified goals [CN02a, Li04]. In addition, both approaches suggest to manage fre-

[1] The Agile Manifesto. `http://www.agilemanifesto.org`

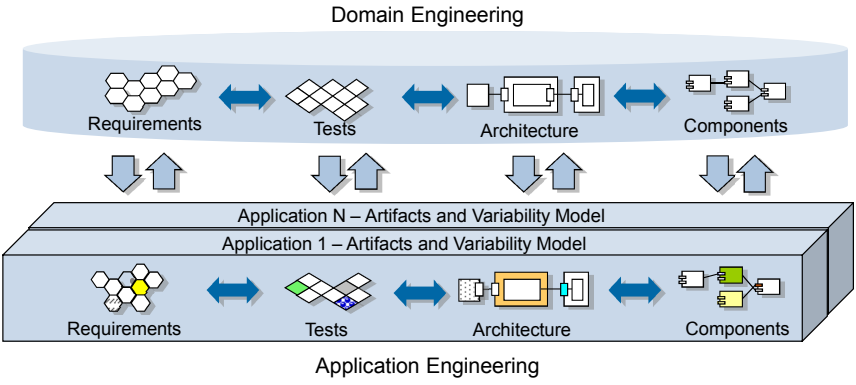


Fig. 1: Domain and application engineering (based on [PBL05]).

quent requirement changes efficiently [TC06]. In the context of APLE, it is tried to establish a hybrid of both approaches. Different researches state conflicts between ASD and SPLE [MRS10, HF08]: the nature of ASD enforces short development cycles with a short planning period, while the establishment of a SPL requires the prediction of upcoming requirements by an intensive requirement analysis phase.

Reducing the long-term investment during the DE phase is one of the main reasons for combining SPLE and ASD. Enforcing the performance of DE before the AE has been identified as very cost-intensive and risky [Dí11]. Establishing and maintaining a SPL leads to significant coordination overhead and slower release cycles [Bo10]. In the context of frequently changing demands, the threat of losing long-term investments becomes even more realistic. For a supplier traditionally dealing with frequent change requests from customers and with having low interest for long-term investments, these aspects are of even higher importance. In addition, [Ca08] highlights the importance of an intensive feedback from AE teams to the DE team. One posibility to avoid long-term integration could be enforcing a short-term integration, which would move the effort to smaller but more frequent system integrations [Bo10].

Very important for an iterative extraction of a product line based on established products is the necessity to identify the most applicable parts of a product for an extraction. Therefore it is inevitable to derive related metrics. In [Be10, BRR10], a set of metrics to measure such an ability is defined. These metrics are based on three levels of component similarity: first, the components' names are compared (extrinsically equal) afterwards their interfaces (syntactically identical) and finally their behavior (semantically identical). While the interface similarity is derived on the basis of ids and signatures only, the behavioral similarity is evaluated by applying given test sequences. In [BRR10], the metrics have been adapted to also measure gradual similarities between components. This kind of relationship model has been applied to identify components which are, to a specific degree, either similar to all other products, just to one or to none. By applying this model, different signatures are matched based on their related output only.

3 Application Engineering focused Software Product Line Development

Recapitulating the statements from Section 2, a SPL development method is required which provides essential feedback for DE during AE, but reduces the additional effort for AE to a minimum. A method which does not require long-term decisions and which keeps the SPL up-to-date and identifies new potential for reusable components is needed.

In the following section, such a method is described which is shown in Figure 2. The coloured activities represent activities which are performed by DE, while the white ones are performed during AE. The main important development artifacts used during this process are the reference architecture, project architecture, component specification, test cases and the component implementation (not shown). In general, the developers from the AE have access to the reference architecture of the SPL, while the developers from DE have

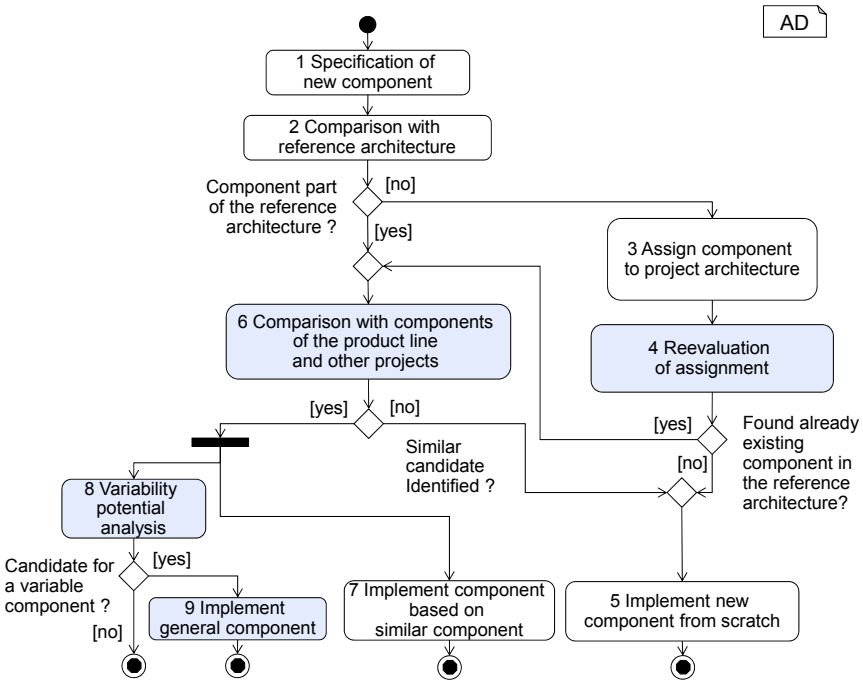


Fig. 2: Method for agile software product line development focused on application domain. (colored = DE activities, white = AE activities)

access to all projects and related development artifacts. In addition, each project has a specific project architecture which is very similar to the reference architecture.

The proposed method follows a component-based AE-first approach: the main item of the software architecture is a component and any specification for a component is first raised during AE. A component will only be considered to be redesigned in a more general way, if it is proven that a corresponding demand is given and a general component is a realistic opportunity within several projects. In addition, AE shall gain most benefit from the already established product line and components which are currently developed in different projects, without being slowed down by the burden of being dependent of generalized components. Next the steps shown in Figure 2 are described in detail, whereby the order defined in Figure 2 is followed.

In the first step, a new component is specified (*1 Specification of new component*). This is done by performing a first draft of the interface with a related functional description. In addition, its position in the software architecture is estimated. This is done by considering the reference architecture of the SPL (*2 Comparison with reference architecture*). In the context of PERSIST, components are hierarchically arranged in a set of compositions. If a suitable component can be identified, which seems to be similar or identical to the specified component, the name and location of the component will be adapted to

the reference architecture. This is the first step, where the SPL supports the development during AE. Complex architectural decisions can often be supported by experiences from the past (which are stored in the reference architecture). In case the component cannot be mapped, a specific position in the project's software architecture needs to be defined (*3 Assign component to project architecture*).

In general, the introduction of new components (components which are not part of the current reference architecture) should be avoided, where possible, to reduce the complexity of the reference architecture. In addition, further evaluations of similarities can only take place efficiently, if components could be mapped to a similar context (name, composition). Therefore, in a second step the decision made in *3 Assign component to project architecture*, is reevaluated by the DE team to avoid any false negatives (*4 Reevaluation of assignment*). Only if both the AE team and DE team cannot map the new component to the reference architecture, a further similarity analysis will not take place. To reduce the effort for AE during step *2 Specification of new component* the main responsibility to avoid false negatives lies with the DE team. If the component cannot be matched the complete implementation will be performed from scratch (*5 Implement new component from scratch*): no benefit can be gained from the SPL but the additional effort for the AE team is also small. Furthermore, the new component is added to the reference architecture after its location has been agreed on (can also be applied after step *4 Reevaluation of assignment*).

If the component specified in the context of the project can be mapped to a component an extrinsical equality [Be10] regarding the reference architecture can be established. This connection cannot only be used to compare the component, which is currently under development, with general components of the SPL but also to compare it with other individual, but extrinsical, equal components from different projects.

During such a comparison (*6 Comparison with components of the product line and other projects*) the most similar component is derived by the DE team by applying a gradual version of the identity relationship model [Be10]:

$$\Delta_{\approx}(c_x, c_y) = \frac{|W|}{|S_{c_x} \cup S_{c_y}|} \qquad \text{with W} = \{s \in S_{c_x} | \exists t \in S_{c_y} : s = t \wedge b_s = b_t\} \tag{1}$$

whereby the syntactical interface set S_{c_j} of component c_j consists of r concrete signatures $s_k = id \times \mathscr{P}(\mathbb{N}, \mathbb{R}, [TYPE], ..\})$ and $b_k : s_k \rightarrow \mathscr{P}(\{\mathbb{N}, \mathbb{R}, [TYPE], ..\})$ prescribes the behavior of the signature s_k [Be10].

To be able to identify semantical similarities automatically, corresponding test cases need to be provided, as suggested in [BRR10]. As consequence a TDD is required to derive similarities as early as possible. If they cannot provided, only structural similarities will be extracted automatically. A semantical analysis has to be performed by experts, contradicting the general goal to reduce the additional effort for the AE team to a minimum. In the best case, the similarity is 100%: the whole component can be reused directly, no additional effort is given. In the worst case, no useful similarity can be identified: the component has to be implemented from scratch (*5 Implement new component from scratch*).

If a candidate with a similarity lower than 100% is identified, this candidate can be used by the AE team to finalize their own specification and the implementation can be based on the provided development artifacts (*7 Implement component based on similar component*). The AE team does not spend any additional effort on implementing a reusable component which is able to fulfil the requirements of both or more similar components. The identified similarities are only used to speed up the development during AE.

In a parallel step (*8 Variability Potential Analysis*) the DE team evaluates if the identified similarities provide enough potential for a general reusable component. Based on the amount of similar components and the degree of similarity on structural and semantical level this decision has to be made (*7 Implement component based on similar component*). If the potential is high enough, the DE team will implement a general component which can then be reused in further projects.

Depending on the actual milestones in the project and the necessary effort to implement the general component, the AE team could also decide to skip their own individual implementation and directly apply the general component. Of course this would further reduce the effort during AE, but in general it is suggested that the necessary time frame or resources are not available. If corresponding expertise and resources are available, the AE team could also implement the general component, therefore performing a DE team task.

The proposed steps ensure that the AE team always specifies component which are as close as possible to the already established reference architecture and related general components. In the best case general components can directly be reused. Additionally all derivations of the reference architecture can automatically be spotted and the degree of variation evaluated. Therefore the potential degeneration of an established product line can be observed continuously and a synchronization between product and product line can be performed iteratively.

4 Evaluation

Since the first application of PERSIST in [RPS14] it has successfully been introduced in 14 different projects (RCP and series development in the context of the automotive power train) . The proposed approach could in part be established at the FEV GmbH. Based on PERSIST which provides a set of standards regarding component and datatype definitions it has already been possible to evaluate potentials and necessary improvements. The described activities are just an extension of PERSIST, no costly adaption of the established process are necessary.

Activities 2, 3, 4 (and of course 1 and 5) are already well established, while for activity 6 and 8 the necessary automatization to efficienty perform these steps is currently missing (see Figure 2). General components, which are used in several projects, have already been established, but always based on intensive manual reviews or during a proactive approach. The possibility to investigate a given reference architecture during the specification of a new component is not noticed as an additional effort. Instead the reference architecture provides helpful information to perform architectural decisions.

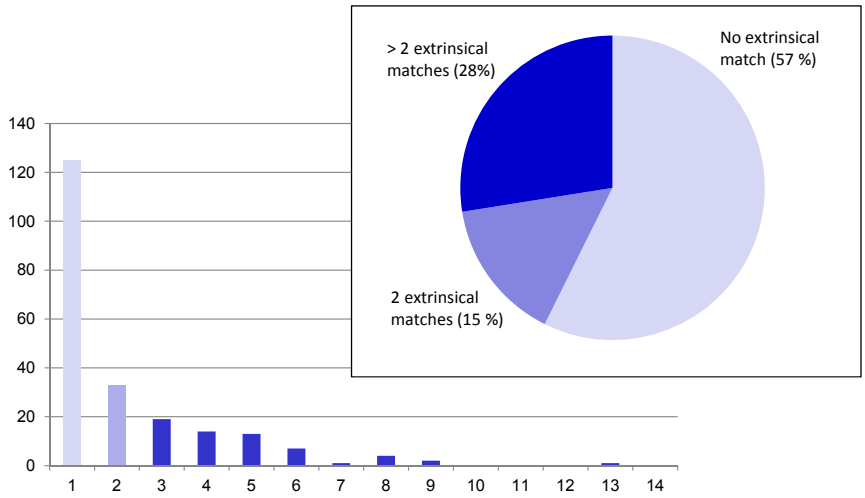


Fig. 3: Amount of components with n extrinsically identical components in several projects.

The comparison made in activity 4 is supported by a nightly build script which compares the different project architectures with the reference architecture. If a component cannot be extrinsically mapped to the reference architecture, a corresponding report will be generated. Based on this report, the DE team can reevaluate the architectural decision and inform the corresponding project, if necessary. This way, the reference architecture can semiautomatically be synchronized. Necessary time-intensive discussions to identify the right location for a new component could not be reduced, but the effort to spot potential deviations has been significantly minimized. Furthermore, it can be evaluated automatically how often a component of the reference architecture occurs in the different projects (how often an extrinsically identical component is given). This information provides the access point for further structural and semantical evaluations (Activity 6). In addition it provides a first rough analysis of potential reuse and a good starting point for every implementation performed by the AE team.

The results of the extrinsical comparison are shown in Figure 3. In the current status the reference architecture consists of 219 components, whereas 125 (57%) components have no extrisincally equal match. 94 (43%) components of the reference architecture are part of more than one project and 61 (28%) of these components are also part of at least three projects. To check the feasibility of the relationship model proposed in [Be10], all components which occur at least in 5 projects have been evaluated regarding their structural similarities. In most of cases, none or very small structural similarities could be identified. A comparison based on the concrete signature, as defined in [Be10], appears not to be sufficient. In many cases, the signatures are not identical but still very similar. During the manual reviews of the components interfaces, many signatures could be identified which only differed in their specific representation: often only different units, similar ranges or

other levels of abstractions result in changed signal names. In addition a manual analysis of different interfaces has been identified as very time-intensive and still error-prone.

5 Related Work

In a reactive manner Salion developed an initial set of systems to derive commonalities and identify variabilities [CN02b] in a second step. This case study already provides first input on a possible collaboration of ASD and SPLE: variable requirements have often been realized by custom solutions instead of generic components [MRS10]. Refactorings have been performed selectively, only smaller adaption on the product line scope, instead of larger changes are performed [MRS10]. In addition, clone detection has been used to identify candidates for an extraction. Still these clone detection mechanisms are not used to identity reuseability potentials during the AE phase.

[Zh11] describes an incremental and iterative reengineering approach which focuses on a component-wise establishment of a product line. This idea is based on a suggestion from [SV02] which proposed to either establish a product line as a whole or component-wise. In [Zh11], based on the developer's expertise, components with a high level of variability and expected intensive maintenance or adaption effort are selected first to be integrated into the SPL. By comparing the maturity level and the degree of similarity of same components in other projects, the reference component can be identified. The approach has been applied in a specific project in which a product line based on 5 products was established. During the incremental establishment of the product line the integration effort and faults report trend significantly decreased, which resulted in positive accumulated profit after the second product had been derived. In comparison to the provided approach no explicit metrics have been defined to measure similarities between components or to analyze the strategical importance for the SPL. Therefore a possible automatization of this task is not discussed. In addition an initial comparison of components through an additional mapping to the reference architecture at the beginning of the development cycle is not part of the process. Potentials of reuseability are only identified due to the reengineering of the software product line. Even component-level automated test-suites are provided to ensure a proper reengineering, these artifacts are not used for any kind of similarity analysis.

In [Ca06] incremental design [BA04] and the planning game [Hu05] are applied to PuLSE-I, a reuse centric application engineering approach, to complement the benefits of SPLE and ASD. In a later step the planning game has been adapted to provide feedback between the domain and the application engineering team [Ca08]. In contrast the proposed approach focuses on efficient similarity analysis in the AE phase due to agile techniques like TDD and continious integration. An introduction of the planning game is currently not considered, but maybe will be applied in the future.

In [GM10, GPM08], a test-driven reactive SPLE approach is described and evaluated. In this approach, new variation points are introduced by using the test artifacts as a starting point. Even though the authors mention in [GPM08] the possibility of comparing acceptance tests to evaluate similarities between given systems and new requirements, this idea is not evaluated further.

Compositional software product lines as described in [Bo10] allow to move all necessary coordination overhead to the architecture level: each team decides on its own which new functionality needs to be implemented and ensures that each new version provides a backward compatibility regarding its interface. While this approach is also architecture-driven, the identification of reuseability potentials is performed individually and is not supported by standardized metrics.

In [Ru12, Ru04] a set of agile methods is proposed to define complex systems based on UML/P, a profile of UML. These methods build a foundation for agile model-driven development with focus on flexibility and customer satisfaction [Ru12].

In [Ho12], a reactive and model-driven approach is suggested, which focuses on the application of reusability in parallel to the project-driven development. Already existing modules are optimized stepwise, as the general introduction of a software product line has been identified as a challenging task.

6 Conclusion

The proposed approach provides the possibility of realizing the project-driven implementation work directly in the project (AE) without loosing the benefits of a corresponding software product line. Each project team can benefit from the established reference architecture to speed up architectural decisions, while the similarity analysis performed during DE can provide additional foundations for the actual implementation. In the long-term, general components can be derived whose reusability potential have been proven through several projects. Furthermore the Activity *4 Reevaluation of assignment* can be fully or semi automated, if adequate information is provided. More complex evaluations for the activities *6 Comparison with components of the product line and other projects* and *8 Variability potential analysis* are currently not automated and require some improvements regarding their granularity to establish a quality similar to a manual review. Hence, a more detailed relationship model and an automated evaluation of semantical similarities based on test cases is planned for the future. Regarding the current status the potential of the suggested approach could not be evaluated totally, but necessary additional modifications are currently developed to be able to perform a more detailed analysis in the future. Moreover, it is necessary to provide the AE team all already established data, such as signals, component definitions or nightly build reports, in an adequate and stable manner to increase the acceptance of the product line. Therefore, it is planned to foster PERSIST with a database supporting signal and model management.

References

[BA04] Beck, Kent; Andres, Cynthia: Extreme Programming Explained: Embrace Change. Addison-Wesley, 2nd edition, November 2004.

[Be02] Beck, Kent: Test-Driven Development: By Example. Addison-Wesley Professional, 1 edition, November 2002.

[Be10] Berger, Christian; Rendel, Holger; Rumpe, Bernhard; Busse, Carsten; Jablonski, Thorsten; Wolf, Fabian: Product Line Metrics for Legacy Software in Practice. In: Proceedings of the 14th International Software Product Line Conference (SPLC 2010). volume 2, 2010.

[Bo10] Bosch, J.: Toward Compositional Software Product Lines. Software, IEEE, 27(3):29–34, May 2010.

[Br06] Broy, Manfred: Challenges in Automotive Software Engineering. In: Proceedings of the 28th International Conference on Software Engineering. ICSE '06, ACM, New York, NY, USA, pp. 33–42, 2006.

[BRR10] Berger, Christian; Rendel, Holger; Rumpe, Bernhard: Measuring the Ability to Form a Product Line from Existing Products. In: Variability Modelling of Software-intensive Systems (VaMos). 2010.

[Ca06] Carbon, Ralf; Lindvall, Mikael; Muthig, Dirk; Costa, Patricia: Integrating Product Line Engineering and Agile Methods: Flexible Design Up-front vs. Incremental Design. In: Workshop on Agile Product Line Engineering. 2006.

[Ca08] Carbon, R.; Knodel, Jens; Muthig, Dirk; Meier, G.: Providing Feedback from Application to Family Engineering - The Product Line Planning Game at the Testo AG. In: Software Product Line Conference, 2008. SPLC '08. 12th International. pp. 180–189, Sept 2008.

[CN02a] Clements, P.; Northrop, L.: Software Product Lines: Practices and Patterns. Addison-Wesley, 2002.

[CN02b] Clements, Paul; Northrop, Linda: Salion, Inc.: A Software Product Line Case Study. Technical Report CMU/SEI-2002-TR-038, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, 2002.

[Dí11] Díaz, Jessica; Pérez, Jennifer; Alarcón, Pedro P.; Garbajosa, Juan: Agile Product Line Engineering - A Systematic Literature Review. Software: Practice and Experience, 41(8):921–941, July 2011.

[GM10] Ghanam, Yaser; Maurer, Frank: Extreme Product Line Engineering - Refactoring for Variability: A Test-Driven Approach. In (Sillitti, Alberto; Martin, Angela; Wang, Xiaofeng; Whitworth, Elizabeth, eds): Agile Processes in Software Engineering and Extreme Programming, volume 48 of Lecture Notes in Business Information Processing, pp. 43–57. Springer Berlin Heidelberg, 2010.

[GPM08] Ghanam, Yaser; Park, Shelly; Maurer, Frank: A Test-Driven Approach to Establishing & Managing Agile Product Lines. In: SPLiT 2008–Fifth International Workshop on Software Product Line Testing. p. 46, 2008.

[HF08] Hanssen, Geir K.; Fígri, Tor E.: Process Fusion: An Industrial Case Study on Agile Software Product Line Engineering. Journal of Systems and Software, 81(6):843–854, June 2008.

[Ho12] Hopp, Christian; Rendel, Holger; Rumpe, Bernhard; Wolf, Fabian: Einführung eines Produktlinienansatzes in die automotive Softwareentwicklung am Beispiel von Steuergerätesoftware. In: Software Engineering 2012: Fachtagung des GI-Fachbereichs Softwaretechnik. volume 198 of LNI, Berlin, Deutschland, pp. 181–192, 2012.

[Hu05] Hunt, John: Agile Software Construction. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[Kr02] Krueger, C.: Eliminating the Adoption Barrier. Software, IEEE, 19(4):29–31, 2002.

[Li04] Lindvall, M.; Muthig, Dirk; Dagnino, A.; Wallin, C.; Stupperich, M.; Kiefer, D.; May, J.; Kahkonen, T.: Agile Software Development in Large Organizations. Computer, 37(12):26–34, Dec 2004.

[MRS10] Mohan, Kannan; Ramesh, Balasubramaniam; Sugumaran, Vijayan: Integrating Software Product Line Engineering and Agile Development. Software, IEEE, 27(3):48–55, 2010.

[PBL05] Pohl, Klaus; Böckle, Günter; Linden, Frank J. van der: Software Product Line Engineering: Foundations, Principles and Techniques. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[Pr07] Pretschner, Alexander; Broy, Manfred; Krüger, Ingolf H.; Stauner, Thomas: Software Engineering for Automotive Systems: A Roadmap. In: 2007 Future of Software Engineering. FOSE '07, IEEE Computer Society, Washington, DC, USA, pp. 55–71, 2007.

[Ri15] Richenhagen, Johannes; Venkitachalam, Hariharan; Schloßer, Axel; Pischinger, Stefan: PERSIST — a scalable software architecture for the control of various automotive hybrid topologies. Technical report, SAE Technical Paper, 2015.

[RPS14] Richenhagen, Johannes; Pischinger, Stefan; Schloßer, Axel: PERSIST—A Flexible and Automatically Verifiable Software Architecture for the Automotive Powertrain. Journal of Electrical Engineering, 2:108–115, 2014.

[Ru04] Rumpe, Bernhard: Agile Modeling with the UML. In (Wirsing, M.; Knapp, A.; Balsamo, S., eds): Proceedings of the Radical Innovations of Software and Systems Engineering in the Future. 9th International Workshop (RISSEF'02). volume 2941 of LNCS, Springer, Venice, Italy, pp. 297–309, October 2004.

[Ru12] Rumpe, Bernhard: Agile Modellierung mit UML: Codegenerierung, Testfälle, Refactoring. Springer Berlin, 2te edition, Juni 2012.

[Sc04] Schwaber, Ken: Agile Project Management with Scrum. Microsoft Press, Redmond, WA, USA, 2004.

[Sc12] Schaefer, Ina; Rabiser, Rick; Clarke, Dave; Bettini, Lorenzo; Benavides, David; Botterweck, Goetz; Pathak, Animesh; Trujillo, Salvador; Villela, Karina: Software Diversity: State of the Art and Perspectives. International Journal on Software Tools for Technology Transfer, 14(5):477–495, 2012.

[SV02] Schmid, Klaus; Verlage, Martin: The Economic Impact of Product Line Adoption and Evolution. IEEE software, 19(4):50–57, 2002.

[TC06] Tian, Kun; Cooper, Kendra: Agile and Software Product Line Methods: Are They So Different. In: 1st International Workshop on Agile Product Line Engineering. 2006.

[VRP15] Venkitachalam, Hariharan; Richenhagen, Johannes; Pischinger, Stefan: A generic control software architecture for Battery Management Systems. Technical report, SAE Technical Paper, 2015.

[Zh11] Zhang, Gang; Shen, Liwei; Peng, Xin; Xing, Zhenchang; Zhao, Wenyun: Incremental and Iterative Reengineering Towards Software Product Line: An Industrial Case Study. In: Software Maintenance (ICSM), 2011 27th IEEE International Conference on. pp. 418–427, Sept 2011.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Qualitätssicherung in der Software-Entwicklung am Beispiel des BMW i3

Peter Schiele[1], Alexander Siller[2]

Abstract: Die E-Mobilität stellt hohe Ansprüche an ihre zugrunde liegende Entwicklung. Schnellere und kosteneffizientere Abläufe sind unumgänglich, um mit der Entwicklung im Wettbewerb Schritt zu halten. Das gilt nicht nur für die HW-Entwicklung des Energiespeichers oder der E-Maschinen, sondern auch für die eingebettete Software des E-Antriebsystems.

Die Entwicklung des BMW i3 wartete mit etlichen neuen und zu Beginn teils unbekannten Herausforderungen auf. Reorganisation im Großen wie auch im Kleinen, Optimierung der Infrastruktur und eine kontinuierliche Prozess- und Produkt-Optimierung – sie alle bildeten die Voraussetzung, um die erwarteten Qualitätsansprüche in einem angespannten Termin- und Kostenrahmen zu erfüllen. Die notwendigen Veränderungen waren der Schlüssel zum Erfolg. Veränderungen beinhalten aber auch das Risiko, Projektziele hinsichtlich Software-Qualität nicht zu erreichen.

Dieser Artikel beschreibt, wie in der Praxis Veränderungsbedarfe zielgerichtet anhand von Metriken ermittelt und gesteuert wurden, aber auch die Metriken selbst Änderungen unterlagen.

Keywords: Metriken, Kontinuierlicher Verbesserungsprozess, Qualitätssicherung, Projektsteuerung, E-Antriebe, Zielvereinbarung

1 Einleitung

Die SW-Qualitätssicherung beim E-Antrieb des i3 setzte auf unterschiedliche Industrie-Standards auf. Diese beschreiben, wie Prozess- und Qualitätsverbesserungen strukturiert umgesetzt werden:

Die ISO/TS 16949 [Is09] gibt Anforderungen vor, deren Erfüllung kontinuierliche Verbesserungen verspricht. Six Sigma [Lz07] liefert einen Werkzeugkasten an Methoden. CMMI [CK11] bietet Best Practices um stufenweise zu reiferen Prozessen zu gelangen. Alle haben eines gemeinsam: Messen und Analyse zählen zu den Kerninhalten, um Projekte zu reifen Prozessen zu führen. Allerdings sagte keine Norm etwas darüber aus, welche Messgrößen sinnvoll angewendet werden können oder sollten.

[1] BMW AG, EA-41, Taunusstr. 41, 80788 München, peter.schiele@bmw.de
[2] BMW AG, EA-41, Taunusstr. 41, 80788 München, alexander.siller@bmw.de

Die Auswahl der Metriken muss sich gemäß Goal/Question/Metric-Methode [SoBe99] an der Frage der Zielsetzung und an den Anforderungen des Entwicklungsprozesses orientieren. Dieser Aspekt wird am Beispiel des BMW i3 im folgenden Abschnitt erläutert.

Der Abschnitt danach geht auf die Definition der Metriken, die Infrastruktur für eine zuverlässige, effiziente und flexible Datenerfassung sowie deren Aufbereitung ein.

Im Anschluss werden die Analyse, die Umsetzung von Maßnahmen und wirksame Steuerungsmechanismen im Projekt i3 dargestellt.

Abschließend wird der Evolutionsverlauf und eine mögliche Standortbestimmung eines beliebigen Messsystems betrachtet.

2 Metriken leiten sich von Zielen ab

Orientierungsrahmen für die Metriken sind die Projektziele. Die Messgrößen bilden den Ausgangspunkt und den aktuellen Start von Verbesserungsprozessen ab. Gleichzeitig stellt sich aber auch die Frage nach der Wirksamkeit von bereits ergriffenen Maßnahmen.

Standardmäßig werden Ziele wie hundertprozentige Testabdeckung, Null Fehler, Reduzierung der Entwicklungskosten o. ä. genannt. Am besten noch alle drei in Kombination.

Auch wenn es den Entwicklerteams oft als nicht realistisch erschien, solche Zielvorgaben unter gegebenen Rahmenbedingungen zu erreichen, so war deren Forderung doch legitim. Wichtig war es bei solchen Metriken, einen zeitlichen Verlauf wie „Burndown“ o.ä. als Zielkurve zu vereinbaren, die realistisch erreichbar war, damit der erste Widerstand gebrochen wurde.

2.1 Ziele aus Projektanforderungen

Testabdeckung: Die Software für den elektrischen Antrieb beschleunigt und bremst das Fahrzeug über die E-Maschine und ist damit Teil der sicherheitsgerichteten Entwicklung. Aus der Anforderung an die funktionale Sicherheit folgte unmittelbar, dass die notwendige Testabdeckung auf Anforderungs-, Design- und Modul-Ebene für die am Sicherheitskonzept beteiligten Software-Einheiten erreicht werden musste. Jede Testlücke in der Software stellte ein Risiko dar. Jeder Fehler musste vermieden werden.

Änderungsaufkommen und Produkt-Stabilität: Der Innovationsgrad in diesem Projekt lieferte einen kontinuierlich wachsenden Ideenpool für Verbesserungen und Änderungsbedarf an den Konzepten. Diese Änderungen brachten ein Risiko mit sich, welches – je näher der SOP rückte – vermieden werden musste. Da die einzelnen Software-Teams im BMW i3 Projekt nach agilen Methoden und daher weitgehend

eigenverantwortlich arbeiteten, waren Metriken zu Änderungsaufkommen und zur Produkt-Stabilität notwendig, um den Scrum-Teams ein Steuerungsmittel an die Hand zu geben.

In Abb. 1 ist der prozentuelle Anteil an geänderten Codezeilen im Vergleich zu den letzten Releases für einzelne Software-Komponenten dargestellt. Die rechte Graphik verdeutlicht dies anhand der Farbgebung von Grün, über Gelb nach Rot. Je mehr die Farbe ins Gelb und dann ins Rot übergeht, desto mehr Änderungen sind eingeflossen. Die Änderungsrate gibt zugleich einen Hinweis, wo die Testschwerpunkte liegen sollten.

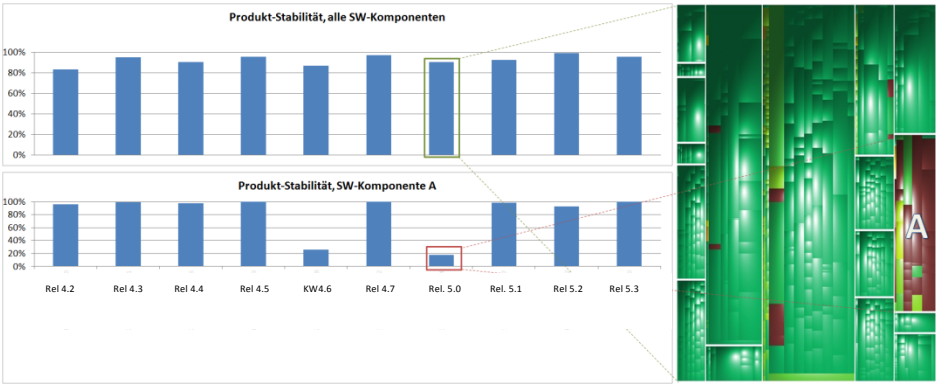


Abb. 1: Beispiel Metrik: Produkt-Stabilität

2.2 Ziele aus Prozessanforderungen

Termintreue: Mittels Meilensteintrendanalyse wurden Planungsrisiken, kritische Pfade, sowie Leistbarkeitsthemen identifiziert, indem der Änderungszeitpunkt des Meilensteins mit dem jeweiligen Zieltermin verglichen wurde. Dies verhalf zu einer realistischeren Terminplanung und entsprechender Termintreue.

Schnittstellenkonsistenz: Anhand von automatischen Prüfungen konnte die Prozessdurchgängigkeit und die Konsistenz zwischen den Prozessschritten (z.B. AUTOSAR-XML) sichergestellt werden. Die Konsistenz war damit frühzeitig hergestellt und ersparte aufwändige Korrekturschleifen bei der Software-Integration. Dabei wurde die Anzahl der Abweichungen zwischen Simulink®-Modell und AUTOSAR-XML gemessen.

Richtlinienkonformität: Umfangreiche Richtlinienchecks auf Funktions-Modell- und auf Code-Ebene verringerten das Fehlerrisiko in der Software, insbesondere das Risiko jener Fehler, die sonst nur durch hohen Testaufwand gefunden werden konnten. Zentrale Messgröße waren die Anzahl der Fehlermeldungen von QA-C.

In- Outflow Problemtickets: Fehler-In- / Outflow dienten als Prognose-Methode für

zukünftige Fehleraufkommen und Abbauraten. Dies führte zu dem Ziel, die geforderte Reife zu den entsprechenden Fahrzeug-Terminen zu erreichen.

2.3 Metriken aus internen Zielen

Durchlaufzeiten: Durch Messung der Geschwindigkeiten bei der Umsetzung und Integration von Änderungen und Fehlerbehebung wurden Langläuferthemen identifiziert und Fehlerabbauraten erhöht.

Wartbarkeit, Änderbarkeit: Komplexitätsmaße gaben ein Indiz für die Wartbarkeit und Änderbarkeit der Software. Neben geläufigen Metriken wie McCabe-Index auf Code-Ebene wurden auch Clone-Checker verwendet, um Redundanzen in der Modell- und Codeentwicklung zu identifizieren.

Effizienz: Die Wiederverwendung von Software für unterschiedliche Fahrzeugvarianten reduzierte den Aufwand in der Entwicklung und in den nachgelagerten Testaktivitäten. Die Wiederverwendung von sogenannten Baukastenelementen wurde durch Codevergleich gemessen. Durch die Einsparung von Codevarianten konnten die Entwicklungsaufwände über die Fahrzeugprojekte hinweg und damit auch im i3 deutlich reduziert werden.

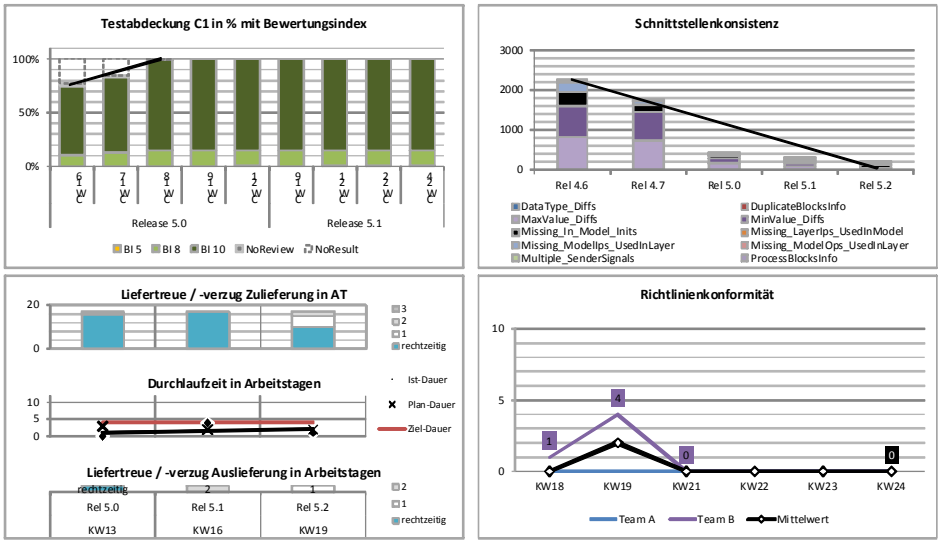


Abb. 2: Metriken für Testabdeckung, Schnittstellenkonsistenz, Durchlaufzeit und Richtlinienkonformität. Die Darstellung der Metriken ist auf Liniendiagramme und gestapelten Balken begrenzt. Auf- und Abbaukurven sowie Ziele geben den Sollverlauf vor.

2.4 Ziele und Metriken als Führungsinstrument

Die Vereinbarung der Ziele erfolgte in den Fachbereichen typischerweise über verschiedene Management-Ebenen, innerhalb der Entwicklungsprojekte in Abstimmung mit den Projekt-Hierarchieebenen oder im Rahmen von aufgesetzten Prozessprojekten.

Als Vorbereitung zur Zielvereinbarung hat es sich bewährt, erste Messungen durchzuführen, um die Erreichbarkeit der Ziele zu gewährleisten. Dazu bedurfte es einer weiteren Verfeinerung in Form der Metrikdefinition, die das Ziel für die Messmethode beschreibt.

Auch im Rahmen des i3-Projektes hat sich gezeigt, dass sich nach einiger Zeit die erwünschten Verbesserungen und Veränderungen im Team-Verhalten nach Identifikation eines Problembereichs und Vereinbarung von entsprechenden Metriken ergeben haben.

3 Wer seine Ergebnisse nicht misst ...

.. der wird seine Ziele nie erreichen. Die Grundlage der Messung bilden die Basisdaten, die Berechnungsformel, der Umfang der Messung und die Bewertung in Form eines Ampelstatus. Die daraus resultierenden Metriken stellen komplexe Sachverhalte einfach und überschaubar dar.

3.1 Vertrauen durch Objektivität

Metriken definieren die Ziele sehr konkret. Weil sie Transparenz herstellen und evtl. Abweichungen vom Ziel quantitativ aufzeigen, sind sie teilweise konkreter als vom Zielenehmer gewünscht.

Damit der Zielenehmer nicht in den Konflikt zwischen objektiver Messung und Zielerreichung kommt, wurde die Aufgabe Metriken zu definieren im i3-Projekt einem unabhängigen Software-Qualitätsteam übertragen. Die Aufgabe beinhaltete auch die Abstimmung mit den beteiligten Fachstellen, das Durchführen der Messung sowie die Darstellung der Metrik. Das erhöhte das Vertrauen in die Messungen und verringerte zusätzliche Belastung der Entwicklungsteams. Durch die objektiven Messungen konnten subjektive Wahrnehmungen korrigiert werden und weiterer Maßnahmenbedarf aufgezeigt werden.

3.2 Nachschärfen von Zielen und Prozessen

Die Metrikdefinitionen lieferten Angaben über die genaue Definition, Identifikation und vor allem die Verantwortung für Gesamt- oder Teilumfänge des ausgewählten Messumfanges, sowie die notwendige Messfrequenz. Sie wurden mit dem Zielenehmer

abgestimmt.

Die abgestimmten Definitionen trugen dazu bei, die Steuerung zu beschleunigen, da die nachher durchgeführten Messungen als Statusbericht direkt und ohne weitere Abstimmung mit den Steuerungsverantwortlichen an höhere Berichtsebenen weitergereicht werden konnte.

Der Status präventiver Qualitätssicherung, insbesondere die Prozesskonformität erlangte dadurch hohe Transparenz und Aktualität, was schnelleres Gegensteuern bei Prozessabweichungen ermöglichte.

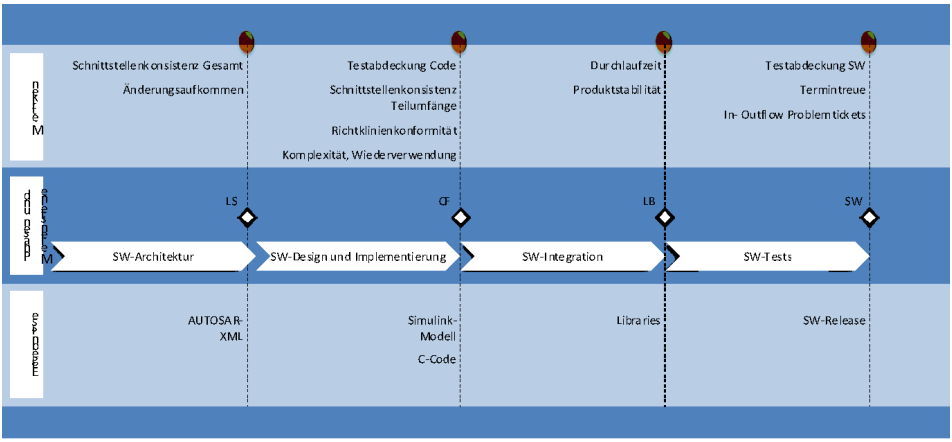


Abb. 3: Metriken werden einzelnen Phasen und Meilensteinen und den zu messenden Ergebnissen zugewiesen.

Die agil arbeitenden Teams blieben im Gesamtentwicklungsprozess eingebunden, weil Qualität und Termintreue gemessen und eingefordert wurden.

3.3 Erste Messung

Die erste Messdatenerhebung bedurf eines tieferen Einblicks in die verwendete IT-Infrastruktur und in die Struktur der abgelegten Daten. Zu den theoretischen Überlegungen erfolgte eine weitere Vertiefung durch die Betrachtung von Sonderfällen. Was gehört zur 100% Zielerreichung und was wird nicht mit gemessen?

Beispielsweise mussten nicht-freigegebene Anforderungen auch nicht getestet werden. Release-Kandidaten oder Entwickler-Software-Stände mussten von Software-Releases unterschieden werden. Verletzungen von Programmierrichtlinien durften nicht als solche gewertet werden, wenn sie der vereinbarten „Deviation Procedure“ unterzogen und in „Whitelists“ dokumentiert wurden.

Durch das richtige Filtern der Daten und die Auswahl der relevanten Informationen stiegt die Signifikanz der Messung und die Akzeptanz der Ergebnisse.

Für die Darstellung der Metrik gilt die einfache Regel: „KISS“ bzw. „Keep it SMART and Simple“. Aus der Darstellung muss intuitiv der aktuelle Status, falls bereits mehrere Messpunkte vorhanden, der aktuelle Trend und vor allem Indizien für den Handlungsbedarf und dessen Verantwortung klar hervorgehen. Das ist im aller Regel ausreichend. Die Rückverfolgbarkeit der Messung ist nach dem gleichen Prinzip über die Metrikdefinition gegeben.

Mit diesem Ansatz konnten die eigenverantwortlichen Teams im i3-Projekt damit nicht nur ihre eigenen Ergebnisse einsehen, sondern auch die anderer Teams. Das förderte den internen Wettbewerb und auch den Austausch. Wo ein Team noch Aufholbedarf hatte, konnte es vom dem anderen Team lernen, das schon am Ziel angekommen war.

Für die Teams war nun klarer verständlich, was die Zielerwartung ist, wo sie aktuell stehen und wo noch Handlungsbedarf existiert. Die Metrik lieferte dafür ein geeignetes Medium zur Kommunikation.

3.4 Metriken automatisieren

Sofern nicht Projektmeilensteine oder Baselines andere Taktraten vorgaben, wurde mindestens wöchentlich gemessen. Die hohe Messfrequenz erforderte eine Automatisierung, zum einen um den Aufwand gering zu halten und zum anderen um die Qualität der Messungen zu gewährleisten und keine Fehler durch manuelle Schritte zu machen. Deshalb bedurfte es für die Erstellung der Metriken einer eigenen Infrastruktur, die in Abb. 4 dargestellt ist.

- Eine automatische Importfunktion sammelte sämtliche Basisdaten aus unterschiedlichen Systemen und Formaten automatisiert ein und legte diese strukturiert, in Form einer SQL-Datenbank, ab. Der Abgleich in die Datenbank erfolgte zentral und skriptgesteuert nach einem fest definierten Messplan, welcher sich aus der Updatefrequenz der zu bedienenden Metriken ableiten ließ.

- Die Datenbankstruktur ließ Zusammenhänge der Daten einfacher erkennen und bot gezielte Abfragemöglichkeiten anhand von SQL-Queries.

- Die Darstellung der Metriken ergab sich durch initial erstellte Diagramm-Formate und dem Update der Datenbankabfrage. Metriken konnten zentral vom Software-Qualitätsteam oder von den einzelnen Steuerungsverantwortlichen aktualisiert werden. In beiden Fällen übernahm das Qualitätsteam die Verantwortung für die Korrektheit der Messung, Datenerfassung und Aufbereitung.

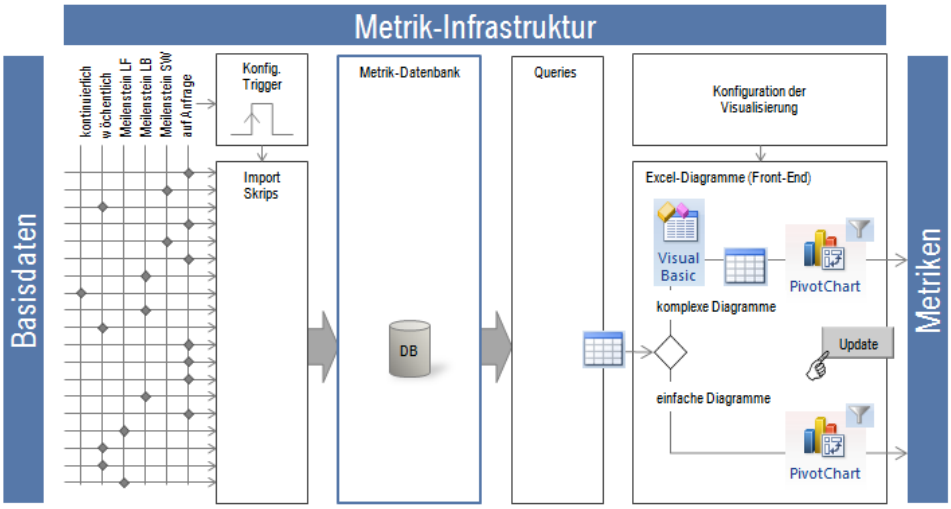


Abb. 4: Prinzipieller Aufbau der Metrik-Infrastruktur: Links der automatische Import der Basisdaten in die Datenbank, rechts die Datenabfrage mit Aggregation und Filterung der Daten in Diagrammen.

Verbesserung der Objektivität wurde durch die Automatisierung erreicht, da sie manuelle Eingriffe unterband.

Die Infrastruktur wurde für das i3-Projekt entwickelt und bediente sich gängiger Technologien für Skripte zur Automatisierung, wie Perl, Datenbanken, wie SQL und Diagrammerstellung mit Excel, welche den Anwendungsfällen gerecht wurden. Die schnelle und flexible Erweiterbarkeit zählte neben dem Verwendungszweck selbst zu einem der wichtigsten Anwendungsfälle und sorgte dafür, dass die Infrastruktur für aktuelle Projekte unverzichtbar geworden ist.

4 Analyse und Steuerung

Eine gut gelungene Metrikdefinition vereinfacht die Analyse, ersetzt sie aber nicht vollständig. Defizite, welche zur Zielabweichung führen, müssen nicht unmittelbar beim Zielenehmer liegen, sondern stammen oft aus vorangelagerten Prozessschritten.

Dies hat sich auch im i3-Projekt am Beispiel Software-Integration gezeigt. Hier wurde die längere Integrationsdauer durch inkonsistente Software-Schnittstellen verursacht. Spätes Bugfixing und Nachliefern von Release-Ständen waren die Folge von vorhandenen Testlücken in den vorangegangenen Prozessschritten.

4.1 Probleme erkennen

Zur Identifikation des Handlungsbedarfes in einer kausalen Kette bewährte sich die Aufteilung von Information auf mehrere Metrik-Diagramme, z.B. Software-Liefertreue und Anzahl der Nachlieferungen. Aus der Analyse ergaben sich weitere zugrunde liegende Ziele, entsprechende Abweichungen und evtl. neue Verantwortliche. Den Verantwortlichen könnten die Metriken individuell zugewiesen werden und versetzten sie in die Lage, selbst zu steuern.

Das bot gleich zwei Vorteile:

- Komplexere Probleme wurden in einfachere zerleget und
- Fehler wurden da behoben, wo sie entstanden.

Das Messsystem wurde somit sukzessive verfeinert und führte zu einem stetigen Zuwachs an Metriken. Damit wurden Defizite unmittelbar dort aufgezeigt, wo die Ursache für Risiken verborgen lag und sie konnten schnell behoben werden. Ziele zur Fehlervermeidung wurden hinzugefügt.

4.2 Metriken strukturieren

Der Zuwachs an Metriken warf neben dem wachsenden Aufwand bei der Metrikerstellung auch das Problem des wachsenden Steuerungsaufwandes inkl. Ursachenanalyse, Erarbeitung von Maßnahmenvorschlägen und deren Abstimmung auf.

Die Automatisierung löste das Aufwandsproblem bei der Erstellung. Der Aufwand für die Steuerung und aller damit verbundenen Abstimmungen blieb noch bestehen.

Sämtliche Metriken wurden deshalb zu steuerungsverantwortlichen Rollen zugewiesen und erhielten dadurch die notwendige Verankerung im Prozess. Das sorgte zusätzlich für mehr Nachhaltigkeit bei der Zielerreichung.

Ein Software-Architekt war nun selbst für die Steuerung mittels der ihm zugeteilten SW-Architekturmetriken verantwortlich. Gleiches galt für die Rolle des SW-Integrationsleiters, des Problemmanagers, des Testmanagers und aller anderen zugewiesenen Rollen.

Im gleichen Zuge wurden die Berichtswege und die Berichtsformate festgelegt, über welche die Metriken weitergeleitet und die Maßnahmen abgestimmt wurden.

Damit hat das i3 Projekt eine Infrastruktur für ein Regelwerk geschaffen, mit dem Veränderungen in Projekten, Prozessen und sogar Organisationen systematisch herbeigeführt werden können.

			Berichtsgremien			
Steuerungs-verantwortlicher	**Metriken**	**Ziele**	**CCB-TPK**	**SW-BKT**	**SW-Prj.-Runde**	**Teststrat.-Runde**
CCB-Leiter	JIRA-Tickets, Anzahl je I-Stufe	Änderungstickets müssen ab bestimmter I-Stufe auf bestimmte %-Anteile reduziert werden.			●	
	Termintreue JIRA-Tickets	Einhaltung Prozessvorgabe: Bestimmter Status muss zu bestimmten Meilenstein erreicht sein.			●	
ECU-Architekt	CPU-Load	Vereinbarte Maximalgrenzen werden nicht überschritten.	●			
	Konformität GS95015	Vollständige Erfüllung der anwendbaren Anforderungen aus	●			
	RAM/ROM (Memory)	Vereinbarte Maximalgrenzen werden nicht überschritten.	●			
FDT-Leiter	Abdeckung Codeinspektion	Code-Inspection wurde für alle Sicherheitsmodule durchgeführt		●		
	Abdeckung Modultests	100% Modell- und Codeabdeckung (Decision-Cov., Statement-Cov., MC-DC, C1)		●		●
	Compiler-Warnungen	0 Warnungen		●		
	Ergebnisse Codeinspection	BI > 6		●		
	Ergebnisse Modultests	BI > 6		●		●
	MISRA	0 Verletzungen		●		
	Modellrichtlinienverletzungen	0 Verletzungen		●		
	Modellschnittstellenprüfung	0 Inkonsistenzen		●		
	Testabdeckung Anforderungen	100% Anforderungsabdeckung		●		
Problemmanager	QC-Fehlerbericht	Anzahl QC- Tickets mit offener TIS <= 7 (Inhouse)				
		TIS muss innerhalb von 10 Tagen festgelegt werden.				
SW-Integ.-Leiter	Anzahl PST-Schienen	100% je Entwicklungs-Generation			●	
	Baseline-Review	Alle Artefakte sind je Baseline vollständig und aktuell			●	
	Durchlaufzeit SW-Build	Reduzierung um 25% in 2015			●	
	Liefertreue-Artefakte	Alle vereinbarten Artefakte werden vom vollständig und rechtzeitig ausgeliefert			●	●
	Maximale Dauer inkr. Build	<= 20min			●	
	Meilensteintrendanalyse	Meilensteine sind mindestens 2 Wochen vor Eintreten stabil			●	
	PST-Nachlieferungen	keine PST-Nachlieferungen (= PST-Lieferungen außerhalb			●	
SW-Integr.-Test	Abdeckung Layertests	100% der spezifizierten Signale inkl. Negativtests	●			●
	Ergebnisse Layertests	BI > 6	●			●
	Testabdeckung Änderungen	100% Änderungsabdeckung		●		●
SW-Sicherh.-Verantw.	Termintreue SW-Freigabeempfehlung	Vereinbarter Termin für SW-Freigabeempfehlung wird			●	
	SW-Sicherheitsaudit	keine Abweichung vom Plan			●	

Abb. 5: Auszug aus verwendeten Zielen und Metriken, sowie deren Zuordnung zu steuerungsverantwortlichen Rollen im Prozess und deren Berichtsgremien

4.3 Berichtsgremien und Steuerung

Veränderungsprogramme sind nur dann erfolgreich, wenn sie vom Management getragen werden. Das gilt auch in dem hier beschriebenen Vorgehen.

In Fall des i3-Projektes wurde das Vorantreiben des Änderungsprozesses von der Projektorganisation unterstützt. Das Metriksystem wurde zum Führungsinstrument. Trends zur Ansteuerung von Zielen und die erreichten Verbesserung haben nicht nur Transparenz und Vertrauen geschaffen, sondern wesentlich zum Erfolg des i3-Projektes beigetragen.

Das Metriksystem ist mittlerweile über das i3-Projekt hinaus in der Entwicklung der E-Antriebe bei BMW etabliert und in der Prozesslandkarte sowie über das Berichtswesen verankert. Unterschiedliche Gremien schätzen die Transparenz und verwenden die Metriken im Berichtswesen sowie zur Projektsteuerung.

5 Evolutionstufen

Der folgende Abschnitt fasst die oben dargestellten Schritte zusammen und gibt die Möglichkeit, den aktuellen Stand bzgl. Steuerbarkeit über Metriken für ein Projekt oder eine Organisation zu bestimmen.

Schritt 1: Einzelne Entwicklungsteams stellen ihre Ergebnisse in Form von Zahlen, Daten, Fakten dar und steuern sich damit selbst. Der Einfluss auf andere Teams ist gering.

Schritt 2: Zielvorgaben von Management werden mit den Zielen und den Metriken der Entwicklungsteams in abgeglichen und abgestimmt. Jedes Team berichtet eigenständig.

Schritt 3: Metriken werden zu Prozessen zugeordnet und zentral verwaltet. Die Metrikerstellung wird automatisiert.

Schritt 4: Bei der Analyse entsteht der Bedarf weiterer Metriken in vorangelagerten Prozessschritten und neuer Zielvorgaben.

Schritt 5: Die Metriken werden über Rollenzuordnung und das Berichtswesen verankert und nachgehalten.

Literaturverzeichnis

[Is09] ISO/TS16949 Qualitätsmanagementsysteme, Besondere Anforderungen bei Anwendung von ISO 9001:2008 für Serien- und Ersatzteil-Produktion in der Automobilindustrie, Dritte Ausgabe, Genf 2009

[CK11] Chrissis, M.B., Konrad, M., CMMI 1.3 für die Entwicklung: Richtlinien für Prozessintegration und Produktverbesserung, Boston 2011

[Lz07] Lunau, S. et.al.: Design for Six Sigma+Lean Toolset: Innovationen erfolgreich realisieren, Frankfurt 2007

[SoBe99] Rini van Solingen, Egon Berghout: Goal/Question/Metric Method, 1999

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Tutorials

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmidt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Remanufacturing in the Information Age

Stefan Schleyer[1]

Abstract: The goal of this tutorial is to provide a compact yet comprehensive introduction to service design in the wind energy business. The tutorial aims at giving the participants the theoretical and practical background in design principles considering the spedific domains the services are designed for. During the life demonstration it is shown how different services can be integrated transparently in a mash up application. The target audience of the tutorial is master and PhD students as well as employees from companies who are interested in gaining an insight in service design.

Keywords: Remanufacturing, Service Design, PREMANUS

1 Introduction

For centuries mankind has attempted to exploit the resources available to it through the reuse of objects and artifacts. However, as wealth of modern man has grown, the financial necessity to utilize the lifetime of these objects has dramatically declined, especially in developed countries with high focus on manufacturing.

The demand for new products also placed undue pressure on the world's resources, creating an ethical imperative to conserve and reuse. Remanufacturing is the process of bringing used products to "like-new" functional state with an equivalent quality assurance.

As this activity provides profitability whilst reducing land-filling and usage of virgin material, it is a financially viable as well as a sustainable business concept.

2 Product Usage and Life Cycle

One of the key issues deterring the uptake of remanufacturing is the information gap which is created when products leave the OEM. The information gap is the result of the lack of data on product usage and its lifecycle.

In general, the product user possesses much greater knowledge regarding a product as he has used it, repaired it, and replaced components. This, in turn, results in the fact that the input to the remanufacturing process is of unknown quality. The lack of reliable

[1] SKF GmbH, Mobile Internet Solutions, Gunnar-Wester-Strasse 12, 97421 Schweinfurt, stefan.schleyer@skf.com

information on product usage and lifecycle leads to missed opportunities with respect to increased economic or environmental impact.

The goal of the EU-funded project PREMANUS is to overcome the asymmetric distribution of information in the End-of-Life (EoL) recovery of products, with a special emphasis on remanufacturing. To achieve this goal, PREMANUS provides an on demand middleware which combines product information and product services within one service oriented architecture. In addition to closing the information gap, the PREMANUS middleware would compute EoL-specific KPIs based on product usage data and make recommendations to its users regarding the viability (in terms of profitability, scope, and time) of remanufacturing a product.

3 Services for the Wind Energy Business

SKF is one of the industrial partners in the PREMANUS consortium and has a large presence within the wind turbine industry. They are a major supplier of full bearing sets for the gearboxes and other parts of the turbine system. SKF also develops and markets a conditional monitoring system which actively supervises and records data from more than thousands of wind turbines worldwide.

Wind turbine gearboxes have not had the operating life-expectancy that was first expected. The typical period before either a major overhaul or replacement is required is between six and ten years, rather than the 15 to 20 years originally intended. This phenomenon has created a demand for repair, remanufacture and upgrading of gearboxes within the wind turbine industry.

Based on the eco system provided by PREMANUS, SKF has developed a prototype application to integrate different kind of infrastructure components like Business Decisions Support Systems (BDSS), Condition Monitoring Systems (CMS), measurement devices etc.

All different components have been aggregated within one mash up application and can be accessed through one common user interface. Based on the application, new remanufacturing services can be offered for the (wind turbine) industry.

4 Summary

Service design is essential in the wind energy business. The design principles have to consider the spedific domains the services are designed for. During the life demonstration it is shown how different services can be integrated transparently in a mash up application

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Angewandte Statistik zur Analyse von Nutzerdaten in der Energiewirtschaft

Detlef E. Schröder

Abstract: Individualisierung ist ein wesentliches Schlagwort der Energiewirschaft und das nicht nur in der Tarifentwicklung, sondern auch in der Prognose und Energieerzeugung. Dazu werden Nutzerdaten erhoben und für die Prognose verwendet. Bei den dabei anfallenden Datenmengen ist eine effiziente Verarbeitung und effektives statistisches Ergebnis entscheidend. Beides soll anhand von Beispielen und den Möglichkeiten der Oracle Datenbank verdeutlicht werden. Die freie Statistische Analyseumgebung R spielt in diesem Zusammenhang eine immer bedeutendere Rolle und soll hier auch zur Anwendung kommen.

Keywords: Prognose, Datenbank, Analysemodelle, Nutzerverhalten, R, Data Mining, Smart Meter, Big Data, Oracle

1 Individualisierung in der Energiewirtschaft

Die Energiewirschaft ist von großen Veränderungen betroffen, die sowohl die Erzeugung, als auch den Verbrauch von Energie betreffen. Die Produktion ist individualisiert, in dem jeder Verbraucher mit einer entsprechenden Photovoltaik, Wind oder Kraft-Wärme-Kopplung Anlage auch zum Produzenten werden kann. Der Markteintritt ist in diesem Beriech sehr einfach geworden und die Hürden geschrumpft. Dem gegenüber ist die Produktion zur Versorgungssicherheit in seiner Komplexität gestiegen.

Auf Grund dieser Effekte wird es für die Energiewirtschaft immer entscheidender auch die Verbraucher Seite zu Individualisieren. Dies kann über die Preis und Vertragsbildung geschehen. Dazu ist es aber notwendig aussagekräftige Modelle zur Verbrauchsprognose zu erstellen und entsprechend schnell zur Anwendung bringen zu können. Die Prognosen Qualität, Ex-ante, lässt sich wesentlich über die erklärenden Variablen steuern und die Funktionsform der Beziehungen [WL24]. Dank den Smart Meter Messwerten und der Big Data Technologie ist die Zahl der erklärenden Variablen enorm gestiegen. Diese mit den in relationaler Struktur vorhandenen Größen in Verbindung zur Anwendung zu bringen stellt allerdings eine Herausforderung dar. Oracle stellt mit der Datenbank, der zur Verfügung stehenden Statistik und der Big Data Ergänzung die Mittel für diese Prognosen zur Verfügung.

2 Statistik mittels SQL

Für Oracle bietet SQL eine verbindende Sprache zur Verwendung sämtlicher Daten. Diese können in der Form eine relationalen Struktur, dem Ergebnis einer R Funktion oder als Big Data Struktur vorliegen. Alle Formate lassen sich mittels SQL verarbeiten und analysieren. Somit stellt SQL die Basis für die Analyse der Nutzerdaten zur Verfügung.

2.1 Prognose-Erstellung

Die Prognose Erstellung folgt den Regeln und dem Ablauf des Data Mining Prozesses. Fragestellung, Daten identifizieren, Datenqualität sicherstellen, Algorithmus identifizieren und Anwenden. Das Anwenden erfolgt über einfaches SQL, das nach folgendem Muster aufgebaut sein kann:

```
SELECT kundennummer, PREDICTION (FOR verbrauch USING *)
OVER() AS verbrauchsprognose FROM alledaten WHERE
bedingungen
```

Im Muster einer Analytischen SQL Funktion wird mit PREDICTION die Data Mining Funktionalität der Datenbank genutzt, um die Vorhersage des verbrauch mit allen Einflussgrößen aus alledaten zu ermitteln. Dabei kann PREDICTION die Oracle SQL eigene Funktion sein, oder eine spezielle R Funktion. alledaten kann dabei relational, Big Data oder das Ergebnis einer anderen R Funktion sein, die Daten als Frame zur Verfügung stellt. Die bedingungen sind dann auf die Daten bezogen und können natürlich in diesem Beispiel vor Allem die Zeitpunkte für die zu berechnenden Verbrauchsmengen betreffen oder natürlich räumlich einschränken.

2.2 Prognose-Anwendung

Aufgrund der einheitlichen Sprache SQL kann jedes SQL fähige Werkzeug diese Prognosen im Dialog verwenden oder die Werte können relational gespeichert werden, pro Kunde ein individuelles Verbrauchsmodell, oder mehrere für unterschiedliche Situationen (tagsüber, abends, Fußballspiel, ...)

Wesentlich ist, das aufgrund der Möglichkeiten der Datenbank keine Einschränkung was Datenmenge oder Performance angeht, vorhanden sind. Somit kann auf ein aufwändiges Sampling verzichtet und die statistische Genauigkeit erhöht werden, da Sampling-Fehler ausgeschlossen sind. Einer Individualisierung steh also nichts im Wege

3 Literaturverzeichnis

[WL24] Wirtschaftslexikon24.com – Stichwort Prognosequalität. Link

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

EnArgus: Zentrales Informationssystem Energieforschungsförderung

Leif Oppermann[1], Elke Hinrichs[2], Ulrich Schade[3], Thomas Koch[4], Manuela Rettweiler[5], Frederike Ohrem[6], Patrick Plötz[7], Carsten Beier[8], Wolfgang Prinz[9]

Abstract: Mit EnArgus stellen wir ein interdisziplinäres Projekt von Energieforschern, Informatikern und Linguisten vor: die Erstellung eines zentralen Informationssystems für die Energieforschungsförderung in Deutschland. EnArgus stellt über ein neues Energieforschungs-Wiki für ausgewählte Themenbereiche der Energieforschung eine Fachontologie bereit, die die Basis für intelligentes Suchen in einer großen Datenbank von Fördervorhaben bildet. Die ontologiebasierte Suchmaschine ist das Kernstück des EnArgus-Informationssystems. EnArgus richtet sich an heterogene Zielgruppen, was sich in verschiedenen Benutzerschnittstellen niederschlägt; wir stellen EnArgus.public, EnArgus.master, EnArgus.wiki und die zugrundeliegende Fachontologie vor und skizzieren den Projektverlauf von der Anforderungsanalyse bis zur Evaluation. Abschließend berichten wir über die derzeit laufenden Weiterentwicklungen im Nachfolgeprojekt EnArgus2.0.

Keywords: Transparenz, Energie, Forschung, Förderung, Wiki, Ontologie, Semantik, Suche

1 EnArgus: Ziele und Anforderungsanalyse

Am Anfang des EnArgus-Projekts stand der Gedanke, staatliche Förderpolitik im Bereich der Energieforschung transparenter zu gestalten und die Bewertung von Technologieentwicklungen zu erleichtern – beides erklärte Ziele des 6. Energieforschungsprogramms der Bundesregierung. Die EnArgus-Partner entwickelten dazu ein zentrales Informationssystem für Energieforschungsvorhaben. EnArgus bietet Wissenschaftlern, Projektträgern und Politikern sowie der interessierten Öffentlichkeit einen einheitlichen und zentralen Zugang zu Informationen über die Energieforschung in der Bundesrepublik. In EnArgus werden keine neuen Daten erhoben, sondern bereits bestehende Vorhabendatenbasen aus verschiedenen Quellen eingebunden. Bisher werden diese Daten meist für die betriebswirtschaftliche Projektbegleitung verwendet. Dagegen geht es in EnArgus® darum, bereits vorhandene Informationen in systematischer Weise

[1] Fraunhofer FIT, Schloss Birlinghoven, 53754 Sankt Augustin, leif.oppermann@fit.fraunhofer.de
[2] OrbiTeam Software GmbH & Co. KG, Endenicher Allee 35, 53121 Bonn, hinrichs@orbiteam.de
[3] Fraunhofer FKIE, Fraunhoferstr. 20, 53343 Wachtberg, ulrich.schade@fkie.fraunhofer.de
[4] OrbiTeam Software GmbH & Co. KG, Endenicher Allee 35, 53121 Bonn, koch@orbiteam.de
[5] Fraunhofer UMSICHT, Osterfelder Str. 3, 46047 Oberhausen, manuela.rettweiler@umsicht.fraunhofer.de
[6] Fraunhofer FKIE, Fraunhoferstr. 20, 53343 Wachtberg, frederike.ohrem@fkie.fraunhofer.de
[7] Fraunhofer ISI, Breslauer Str. 48, 76139 Karlsruhe, patrick.ploetz@isi.fraunhofer.de
[8] Fraunhofer UMSICHT, Osterfelder Str. 3, 46047 Oberhausen, carsten.beier@umsicht.fraunhofer.de
[9] Fraunhofer FIT, Schloss Birlinghoven, 53754 Sankt Augustin, wolfgang.prinz@fit.fraunhofer.de

zusammenzustellen und kontinuierlich an zentraler Stelle bereitzustellen, um die Energieforschungsförderung besser auswerten zu können.

Zu Beginn des Projektes wurde eine ausführliche Anforderungsanalyse durchgeführt. Zu diesem Zweck wurden zunächst durch Befragungen die Anforderungen unterschiedlicher Nutzer (Forscher, Projektträger, Öffentlichkeit) an Projektdatenbanken eruiert. Typische Fragestellungen der Anwender wurden in detaillierter Form zusammengestellt (z. B. Informationslieferung für den Bundesbericht Energieforschung, Erstellung von IEA[10] - Statistiken, Beantwortung von Anfragen aus dem Bundestag, Recherchen von Forschern zu Stand der Technik und Forschung sowie zu möglichen Kooperationspartnern). In diesem Zusammenhang wurde auch eine Auflistung der Probleme erstellt, die bei der Beantwortung dieser typischen Fragestellungen üblicherweise auftreten. Auch die üblichen Vorgehensweisen von Energieforschern bei Recherchen wurden im Rahmen der Anforderungsanalyse dargestellt. In einem nächsten Schritt erfolgte die Ableitung von Anforderungen an das Gesamtsystem EnArgus aus den eruierten Anforderungen von Anwendern und den analysierten Schwächen, aufgeschlüsselt nach sprachlichen Anforderungen, technischen Anforderungen, Anforderungen für eine Weiterentwicklung zur Kooperationsplattform für Energieforscher sowie fachlichen Anforderungen. Anschließend wurden diese Anforderungen von den verschiedenen Nutzern priorisiert. Die Priorisierungen wurden zusammengefasst, ihre Umsetzbarkeit von den IT-Partnern bewertet und so eine finale Priorisierungsliste zur tatsächlichen Umsetzung der Anforderungen im Rahmen des Projektes erstellt.

Darüber hinaus wurde von den Energieforschern anhand von exemplarischen Beispielen aus den im Rahmen des Projektes bearbeiteten Fachthemen eine herkömmliche Auswertung existierender Datenbanken vorgenommen und diese durch die Auswertung weiterer Informationsquellen (z. B. Literatur, Patentdatenbanken etc.) ergänzt. Diese Musterrecherchen dienen im Verlauf des EnArgus-Projektes als Grundlage für den Vergleich der Recherche in konventionellen Datenbanken gegenüber der Recherche mit dem neuen EnArgus-System.

Das Projekt wird vom Bundesministerium für Wirtschaft und Energie im Rahmen des Energieforschungsprogramms gefördert und vom Projektträger Jülich betreut.

2 Das EnArgus-Informationssystem

Den verschiedenen Benutzerkreisen wie Projektträgern, Energieforschern, Ministerien und Öffentlichkeit präsentiert sich das EnArgus-System auf unterschiedliche Weise. Dieser Abschnitt bietet einen Überblick über die wichtigsten Benutzerschnittstellen, sowie die semantische Datenaufbereitung als Grundlage für die Suchfunktion.

Kern von EnArgus ist die Fachontologie, die semi-automatisch aus redaktionell

[10] IEA = Internationale Energie-Agentur

erstellten Wiki-Texten abgeleitet wird. Für die breite Öffentlichkeit bietet EnArgus.public eine ansprechende und leicht zu bedienende Schnittstelle mit einfachen Suchmöglichkeiten. Energieforschern und Projektträgern stellt EnArgus.master eine Kooperationsplattform mit vollem Zugriff auf die Funktionalität eines Recherche-Cockpits zur Verfügung. Der Zugriff auf das EnArgus.wiki ist beiden Benutzerkreisen möglich. Alle Benutzerschnittstellen sind webbasiert, ein aktueller Web-Browser genügt.

2.1 EnArgus.wiki – Das Energieforschungs-Wiki

Das Energieforschungs-Wiki ist Teil des EnArgus-Projektes und wird von den Energieexperten aufgebaut, indem diese zu ihrem jeweiligen Fachbereich Wikipedia-ähnliche Texte erstellen. Mit der Erstellung des Wikis sind zwei Ziele verbunden. Zum einen bietet das Wiki für das EnArgus-System eine Art Nachschlagewerk, in welchem Nutzer die Begriffe, die ihnen weniger geläufig sind, nachschlagen können. Zum anderen dient das Wiki als textuelle Basis zur teilautomatischen Vervollständigung der Energieforschungsontologie.

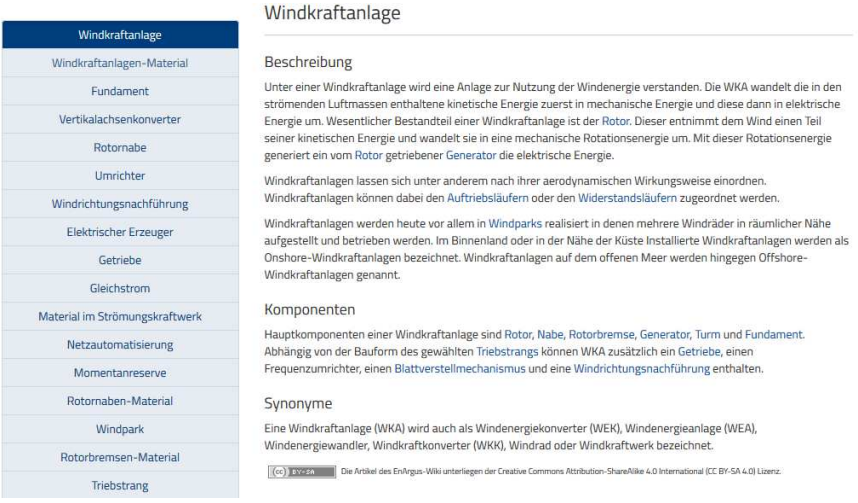


Abb. 1: Wiki-Text zum Thema „Windkraftanlage“, der nach den vorgegebenen Restriktionen erstellt wurde

Der Aufbau des gesamten Wikis ähnelt stark der Wikipedia, da zu jedem relevanten Begriff eine eigene Wiki-Seite angelegt wird (s. Abbildung 1). Diese Seiten werden über zusammenhängende Begriffe verlinkt, so dass eine einfache und schnelle Navigation durch die verwandten Fachbegriffe möglich ist. Das Wiki soll also als ein glossarähnliches Dokument betrachtet und verwendet werden können. Damit das Wiki seine beiden Aufgaben erfüllen kann, müssen die Experten beim Aufbau der Wiki-Seiten und bei den in den Texten verwendeten Satzstrukturen bestimmte Regeln berücksichtigen. So sieht der Seitenaufbau vor, dass der zu definierende Begriff als

Überschrift genutzt wird und stets im Singular steht. Danach folgt die Definition sowie, je nach Begriff, die Beschreibung von Prozessen oder weiteren Komponenten. Im Anschluss können Tabellen eingefügt oder Synonyme des behandelten Begriffes aufgelistet werden.

Im Rahmen von EnArgus wurden zunächst Wiki-Texte zu folgenden Themen erstellt: *Elektrische Energiespeicher*, *Fossil befeuerte Kraftwerke*, *Energieeffizienz in der Industrie*, *Erneuerbare Energien*. Im Rahmen des Nachfolgeprojektes EnArgus2.0, das im Juli 2013 gestartet ist, werden etliche weitere Themenbereiche aus der Energieforschung erschlossen und aufbereitet: *Kompression, Druckluft, Pumpen*, *Prozesstechnologien*, *Elektromobilität*, *Materialforschung*, *Erzeugung von Wärme und Kälte in Gebäuden*, *Verteilung und Speicherung von Wärme und Kälte*, *Energieeffizienz in Städten*, *Biogas*, *Industrielle Stoffkreisläufe (Biokraftstoffe, Bioenergie, energetische Nutzung von Abfällen)*, *Werkstoffe, Konstruktion und Betrieb von Bauten und Anlagen*, *Netze und Stromversorgung*, *Energiewirtschaft und Stromhandel, Energieumwandlung, Energieeffizienz, Systemorientierte, sektorübergreifende Energieforschung, Gesellschaftsverträgliche Transformation des Energiesystems, Rahmenbedingungen*.

2.2 Vom Wiki zur Ontologie

Unter einer Ontologie versteht man die Repräsentation von Wissen in einer Form, bei der Systeme auf dieses Wissen zugreifen und es für ihre Aufgabenstellung nutzen können. Im EnArgus-Projekt stellt die eingebundene Ontologie Wissen über die Energieforschung zur Verfügung, so dass bei einer Suche, die über einen einzugebenden Begriff erfolgt, weitere Begriffe, die mit dem eingegebenen Begriff inhaltlich in Beziehung stehen, genutzt werden können. So verfügen die meisten Begriffe etwa über Synonyme, also über alternative Bezeichnungen, die in der Ontologie als weitere „Label“ zu dem Begriff abgelegt sind. Bei einer Suche mit dem Begriff „Batterie“ kann damit die Suche so erweitert werden, dass auch die (quasi-)synonymen Begriffe – im Fall von „Batterie“ sind dies die Begriffe „Akku“, „Akkumulator“, „Primärzelle“ und „elektrochemischer Energiespeicher“ – für die Suche verwendet werden.

Eine weitere einfache semantische Relation ist die Hyponymie, mit der Über- und Unterbegriffe gefasst werden. Diese Relation lässt sich nutzen, wenn zu einem Teilbereich der Energieforschung Unterbereiche bestimmt werden. Damit können bei einer Frage zur Forschung in einem bestimmten Bereich auch dessen Unterbereiche berücksichtigt werden. In Abbildung 2 sind der übergeordnete Begriff zum Beispielsbegriff „Batterie“ („Energiespeicher“) sowie die untergeordneten Begriffe („externe Speicher“, „Hybrid-Flow-Batterie“ etc.) in dem Begriffsbaum, der sogenannten Taxonomie, im linken Fenster zu erkennen.

Der manuelle Aufbau einer Ontologie ist stets aufwändig und erfordert sowohl Kompetenzen im Bereich der Informationstechnik als auch in dem zu repräsentierenden Gegenstandsbereich, im Fall von EnArgus also im Bereich der Energieforschung. Generell wurde die Ontologie, die im Projekt entstanden ist und die bei den Suchen

genutzt wird, dadurch erstellt, dass die im Projekt arbeitenden Experten zur Energieforschung das bereits erwähnte Energieforschungs-Wiki erstellt haben. Aus den einzelnen Texten dieses Wikis wurden dann die einzelnen Einträge für die Ontologie nach informationstheoretischen Gesichtspunkten extrahiert und in die Ontologie eingefügt. Die so resultierende Fachontologie zur Energieforschung wurde von den Energieforschungsexperten evaluiert, um sicherzustellen, dass die Ontologie keine sachlichen Fehler enthält.

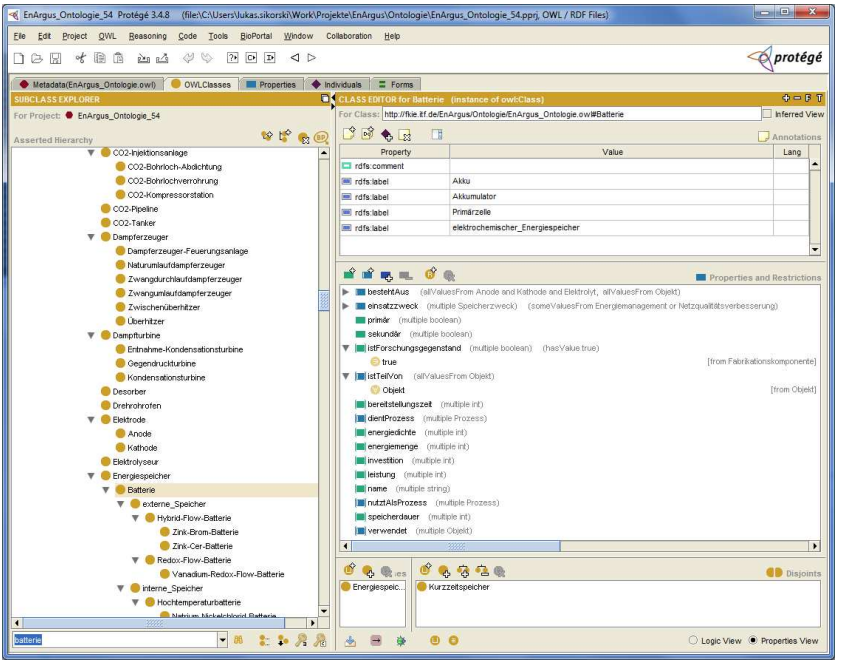


Abb. 2: Snapshot der Fachontologie in Protégé; zu sehen sind die Eintragungen zu „Batterie“.

Um die Erstellung und die Pflege einer Ontologie zu vereinfachen, wurden im EnArgus-Projekt auch Verfahren entwickelt, mit deren Hilfe Ontologie-Einträge teilautomatisch aus den Wiki-Texten entnommen werden können. Der Prozess der teilautomatisierten Gewinnung von Ontologie-Einträgen aus Texten beginnt mit der so genannten „Informationsextraktion“. Dabei wird der Wiki-Text einer Kette von Teilprozessen unterworfen. Auf einen Tokenizer zum Erkennen von Wortgrenzen und Satzzeichen folgen ein Abgleich der erkannten Worttoken mit Gazetteer-Listen, ein Sentence Splitter zum Segmentieren der Sätze und ein Part-of-Speech-Tagger zur Zuweisung der Wortart, der die Grundform generiert und die morphologischen Eigenschaften der Wörter erkennt. Das letzte Modul der Prozesskette ist der General Transducer, der zum Erkennen von (komplexen) Eigennamen (Named-Entity-Recognizer) und syntaktischen Strukturen eingesetzt wird. Dabei werden komplexere Ausdrücke mit der Hilfe eines regelbasierten Pattern-Matching erkannt. Alle Informationsextraktionsmodule annotieren

den Text mit den ermittelten linguistischen Eigenschaften der Wörter bzw. der Sätze. Auf der Grundlage dieser linguistischen Annotationen und unter Hinzunahme von syntaktischen und semantischen Informationen über Verben können die Konstituenten, also die Satzteile der Textsätze, dann semantisch annotiert werden. Das nennt man „Semantic Role Labeling“. Die semantischen Annotationen werden schließlich durch ein abschließendes Modul ausgewertet, welches nach spezifischen Konstellationen (von Wörtern und semantischen Annotationen) sucht, die auf Sätze hinweisen, mit denen ontologisches Wissen sprachlich ausgedrückt wird. Solches entsprechend in den Texten festgestellte Wissen kann dann in die Ontologie übertragen werden.

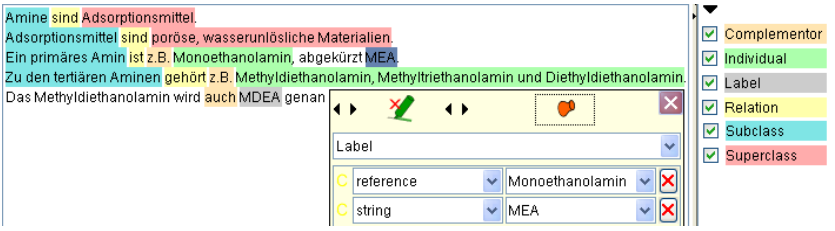


Abb. 3: Beispieltext zum Annotieren von ontologischem Wissen

Abbildung 3 zeigt die semantische Annotation von fünf Beispielsätzen, wie sie im Wiki zu finden sind. Hervorgehoben ist dabei im dritten Satz „MEA“, das als alternative Bezeichnung von „Monoethanolamin“ erkannt und entsprechend annotiert wird. In der Fachontologie findet sich dann allerdings „Monoethanolamin“ als Label zu „MEA“ wieder, weil die angesprochene Evaluation der Ontologie durch die Energieforscher an dieser Stelle dazu geführt hat, dass „MEA“ „Monoethanolamin“ im Begriffsbaum ersetzt hat. Damit wird „Monoethanolamin“ als Label der Klasse „MEA“ geführt. Diese Umorganisation hat jedoch auf die Suchen, in denen als Begriffe „Monoethanolamin“ oder „MEA“ verwendet werden, keinen Einfluss, da die beiden Begriffe jeweils als Bezeichnungsalternativen verstanden werden.

2.3 Effizientes Suchen mit der EnArgus-Ontologie

Bei der Auswertung der EnArgus-Informationsbasis spielen effektive und effiziente Suchverfahren eine wichtige Rolle. Sie sollen es Fachleuten und Laien gleichermaßen erlauben, sich schnell einen Überblick für laufende und abgeschlossene Forschungsvorhaben eines bestimmten Energieforschungsbereichs zu verschaffen und die Ergebnisse auswerten zu können. Intelligente Suchverfahren zur Suche nach semantisch ähnlichen Suchbegriffen verbessern dabei die Ergebnisse von Recherchen im Informationssystem. Die Ontologie spielt dabei eine wesentliche Rolle, indem semantisch verwandte Begriffe für erweiterte Suchanfragen vom System angeboten werden können (s. Abbildung 4).

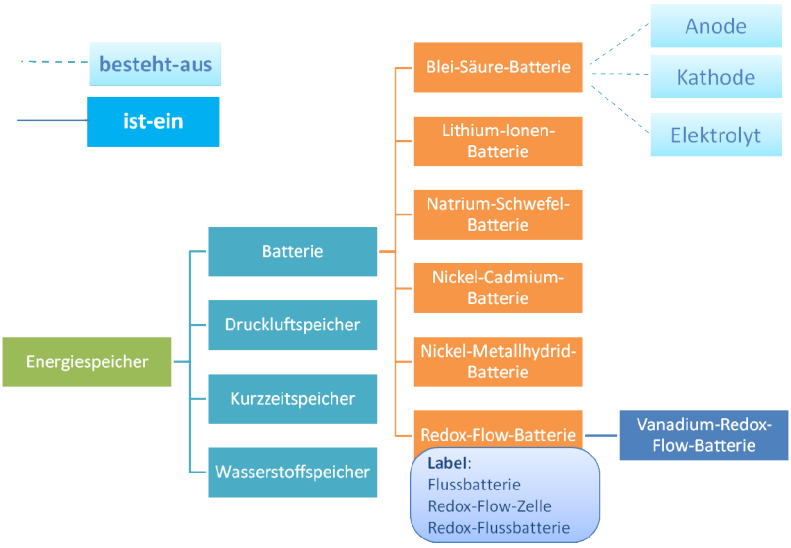


Abb. 4: Auszug aus der Fachontologie. Über das synonyme Label „Flussbatterie“ werden beispielsweise auch Ergebnisse zur „Redox-Flow-Batterie“ gefunden

Der Aufbau der EnArgus-Ontologie erfolgte in Kooperation von Experten aus der Energieforschung und Informatikern. Die Energieexperten verfassten Wiki-Fachbeiträge, die vom System syntaktisch analysiert werden. Daraus werden teilautomatisch Energiefachbegriffe extrahiert, Relationen zwischen Begriffen identifiziert und in die Ontologie gespeist. Die so resultierende Fachontologie zur Energieforschung wurde von den Energieforschungsexperten wiederum evaluiert, um sicherzustellen, dass sie keine sachlichen Fehler enthält. Im Mai 2013, kurz vor Abschluss des ersten EnArgus-Projekts, enthielt die im Projekt entwickelte Fachontologie 710 Klassen, mit denen 1490 Begriffe repräsentiert werden, die über 18 semantische Relationen zueinander in Beziehung gesetzt werden. Zwei Jahre später ist die Ontologie auf bereits über 2300 Klassen gewachsen. Um die Erstellung und die Pflege einer Ontologie zu vereinfachen, wurden im EnArgus-Projekt auch Verfahren entwickelt, mit deren Hilfe Ontologie-Einträge teilautomatisch aus den Wiki-Texten entnommen werden können [Ha12].

2.4 EnArgus.master – Das Vollsystem

EnArgus.master bietet den Vollzugriff auf alle Such- und Recherche-Funktionen für registrierte Benutzer, vor allem bei Projektträgern und Ministerien. Grundlage für die Entwicklung war das von OrbiTeam und Fraunhofer FIT entwickelte BSCW Shared-Workspace-System[11]. Das BSCW-System ist ein webbasiertes Groupware-Werkzeug für effiziente Teamarbeit, das es den Mitgliedern von Arbeitsgruppen ermöglicht, sich in

[11] http://www.bscw.de/

gemeinsamen Arbeitsbereichen auszutauschen. Das BSCW-System bietet neben zahlreichen Grundfunktionen wie Nutzerverwaltung, Zugriffsrechten und Benachrichtigungsfunktionen auch die notwendige technische Infrastruktur zur Entwicklung komplexer, webbasierter Lösungen für spezifische Problemstellungen.

Zur Entwicklung des EnArgus-Informationssystems waren dennoch an verschiedenen Stellen Anpassungen und Erweiterungen der BSCW-Software notwendig. Eine erste Anforderung an das EnArgus-System bestand darin, Daten über Energieforschungsvorhaben aus unterschiedlichen externen Datenquellen zu importieren, zu verwalten, darzustellen und durchsuchbar zu machen. Dazu wurde ein flexibles Metadatenkonzept für Vorhabendaten entworfen und implementiert. Auch können Vorhaben-Objekte mit später hinzugefügten Projektergebnissen (Abschlussberichte, Veröffentlichungen, Projektdokumentation etc.) verknüpft werden. Neben Schnittstellen zum Import von Daten aus externen Datenquellen war dazu eine Datenstruktur notwendig, die eine persistente Verwaltung der Daten zu den Vorhaben ermöglicht. Für die intelligente, semantische Suche wurde die EnArgus-Fachontologie angebunden.

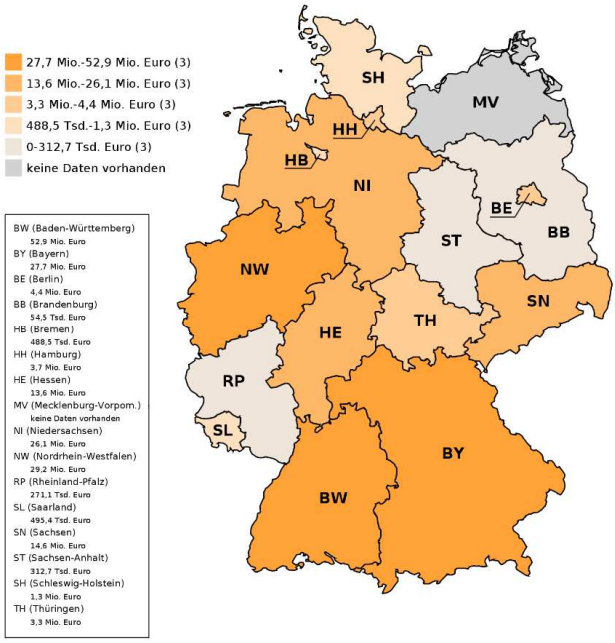


Abb. 5: Geografische Visualisierung einer Suchanfrage verteilt nach Bundesländern

Zur Auswertung der erlangten Suchergebnisse wurde EnArgus mit einem Werkzeugkasten zur Aufbereitung, Auswertung und Visualisierung von Suchergebnissen ausgestattet. Mit diesem "Recherche-Cockpit" können komplexe Suchanfragen

kombiniert, verschiedene mathematische Auswertungsfunktionen durchgeführt und Ergebnis der Recherche in Graphen, Diagrammen oder geografischen Landkarten visualisiert werden. So kann auf geografischen Schwerpunktkarten beispielsweise angezeigt werden, wo in der Bundesrepublik (in welchen Bundesländern oder in welchen Kreisen) besonders viel zu einem bestimmten Energieforschungsthema geforscht wird (s. Abbildung 5).

Bei der Formulierung von Suchanfragen können für diverse Metadaten Bedingungen formuliert werden; für textuelle Daten ist eine Verwendung ungenauer Werte („Wildcards“) als Bedingung möglich. Rechtschreibfehler werden innerhalb gewisser Grenzen toleriert; dazu wurde auf der Basis standardisierter Wörterbücher eine Sprachenerkennung sowie eine Rechtschreibprüfung samt Korrektur auf der Eingabe der Benutzer realisiert. Über die eingebaute Fachontologie unterbreitet EnArgus bei Eingabe von Fachbegriffen darüber hinaus Vorschläge für verwandte Fachbegriffe, wie zum Beispiel Synonyme oder Ober- bzw. Unterbegriffe, nach denen dann ebenfalls gesucht werden kann.

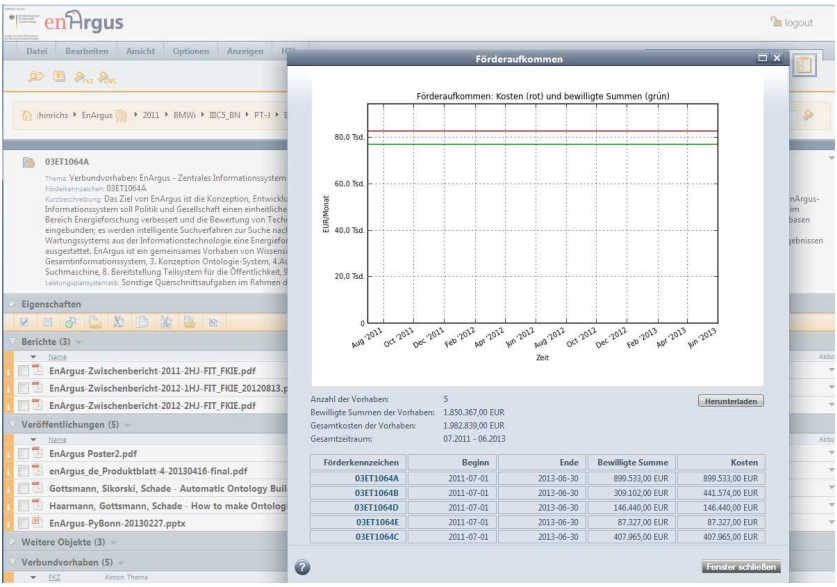


Abb. 6: Ansicht eines Energieforschungs-Vorhabens (hier: EnArgus)

Um die Übersicht zu steigern, können einzelne, ungewollte Vorhaben aus dem Ergebnis ausgeschlossen oder die Ergebnismenge durch Verfeinerung der Suchanfrage oder Anwendung von Filtern („Facetten-Browsing“) schrittweise weiter eingeschränkt werden. Auswertungen können die Ergebnisse mehrerer Suchanfragen berücksichtigen. Im Falle mehrerer Ergebnismengen können diese wahlweise vereinigt oder als separate Gruppen von Forschungsvorhaben berücksichtigt werden. Gruppierungsfunktionen erlauben auch bei einzelnen Suchanfragen eine Gruppierung von Forschungsvorhaben

basierend auf deren bekannten Eigenschaften, wie beispielsweise Kalenderjahr oder Zuwendungsempfänger. Abbildung 6 zeigt die verfügbaren Daten eines Energieforschungsvorhabens (in diesem Fall EnArgus selbst): Vorhabenthema, Angaben zum Verlauf und andere administrative Metadaten sowie das Förderaufkommen. Man erhält direkten Zugriff auf relevante Vorhabendokumente wie Berichte und Veröffentlichungen sowie Referenzen auf Verbundvorhaben und „verwandte Vorhaben", die dem Vorhaben in Beschreibung und Dokumentation semantisch ähnlich sind.

2.5 EnArgus.public – Zugang für die breite Öffentlichkeit

EnArgus.public ist die öffentliche Schnittstelle von EnArgus[12]. Sie bietet auch dem nicht registrierten Benutzer Zugriff zu öffentlich bereitstehenden Daten. Nicht alle Vorhaben in der EnArgus-Datenbank sind allerdings frei verfügbar und nicht alle Metadaten eines Vorhabens dürfen von der Öffentlichkeit gelesen werden, z. B. Kontaktdaten von Ansprechpartnern. Der Zugriff ist daher entsprechend eingeschränkt.



Abb. 7: Ontologie-basierte Suche in EnArgus.public

[12] EnArgus.public ist seit April 2015 unter http://www.enargus.de erreichbar.

Für die zur Verfügung stehenden Daten bietet EnArgus.public eine vereinfachte Bedienung. Die Suche erfolgt in erster Linie über Texteingabe und wird durch Ontologie-basierte Vorschläge ergänzt (s. Abbildung 7). Darüber hinaus kann der Anwender die Suchergebnisse über Filter, sog. Facetten, z. B. Zeiträume oder Fördersummen, eingrenzen (s. Abbildung 8).

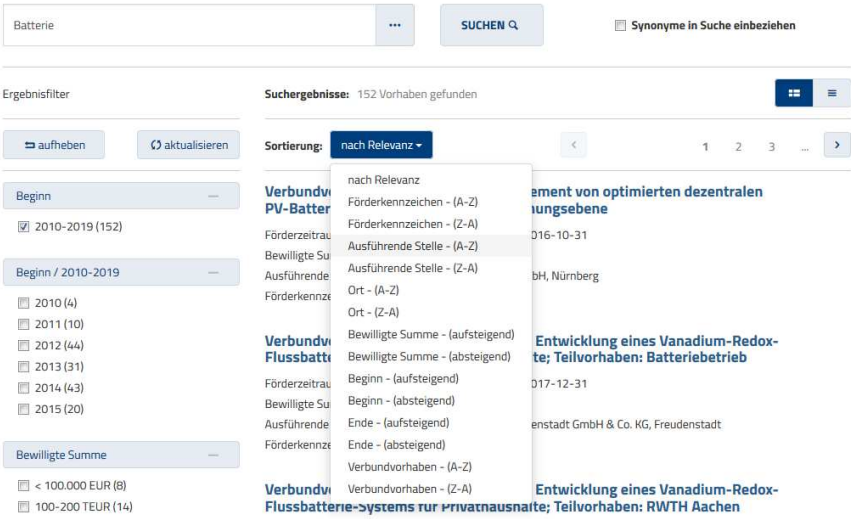


Abb. 8: Ergebnisansicht mit Filter- und Sortierfunktionen

Des Weiteren hat der nicht registrierte Benutzer Zugriff auf die EnArgus-Ontologie, auf EnArgus.wiki, sowie auf veröffentlichte Auswertungen zu Themen wie: *Energieeffizienz in Gebäuden, Quartieren und Städten*, *Energieeffizienz in Industrie, Gewerbe, Handel und Dienstleistungen*, *Brennstoffzellen und Wasserstoff*, *Solarthermische Kraftwerke*.

3 Evaluation des Systems

Das Ziel von EnArgus ist die Konzeption, Entwicklung und Erprobung eines zentralen Informationssystems für Energieforschungsvorhaben aus dem Bereich der Förderung durch Bund und Länder. Zentrale Fragen für die Bewertung des Gesamtsystems sind daher, (1) ob das neu entwickelte EnArgus-System nennenswert bessere Ergebnisse als bisherige Suchmöglichkeiten bietet und (2) wie gut die entwickelte Ontologie die betrachteten Energieforschungsthemen abdeckt. Beispielhaft für die Verbesserung der Suchergebnisse sei hier der Vergleich einer Projektsuche in PROFI[13] und EnArgus für den Bereich Windenergie gezeigt. Einen großen Schwachpunkt bisheriger Datenbanken

[13] PROFI: Verwaltungsinterne Förderdatenbank des Bundesministeriums für Bildung und Forschung, die auch von den Projektträgern verwendet wird. Die öffentlich verfügbare Webschnittstelle FöKat (Förderkatalog der Bundesregierung) greift auf Daten aus der PROFI-Datenbank zu.

stellt die Auswahl der Suchbegriffe bei einer Recherche für ein Themengebiet oder einer Komponente dar. So führte die Suche nach synonymen Begriffen wie „Windenergieanlage“ und „Windenergiekonverter“ zu teilweise deutlichen Abweichungen. Da mithilfe der ontologiebasierten EnArgus-Software dies verbessert werden soll, werden die gewählten Suchbegriffe in der EnArgus-Software einmal mit und einmal ohne Unterstützung der Ontologie eingegeben und mit den Suchergebnissen aus den Musterrecherchen verglichen. Tabelle 1 zeigt die Suchanfragen aus der PROFI-Recherche und aus der EnArgus-Software.

Suchbegriff	PROFI	EnArgus.public		EnArgus.master	
		ohne Synonyme	mit Synonymen	ohne Ontologie	mit Ontologie
Windenergie	54	1781	1781	1781	1781
Windkraft	6	230	230	230	230
Windenergieanlage	1522	1550	1775	1550	2452
Windkraftanlage	99	147	1778	147	3891
Windenergiekonverter	19	19	1775	19	2452
Windkraftwerk	1	9	1775	9	2452

Tab. 1: Suchanfragen in PROFI aus der Musterrecherche und aus der EnArgus-Software

Ähnliche Musterrecherchen mit verschiedenen Zielen und Schwerpunkten lassen folgende Punkte zur Recherche mit EnArgus ableiten:

- Die in der Ontologie abgelegten Synonyme können automatisch in die Suche eingebunden werden und verbessern damit deutlich die Anzahl der Treffer.
- Die Vorschläge der Ontologie helfen sowohl dem Fachfremden bei einer erweiterten Suche (durch die Vorschläge der Ontologie) als auch dem Experten durch die Einbeziehung von semantisch relatierten Begriffen.
- EnArgus kann nur so gut sein wie die zur Verfügung stehenden hinterlegten Daten. Eine wirklich breite Abdeckung aller Energieforschungsprojekte kann nur durch die Erweiterung der Datenbasis gelingen.

Zusammenfassend geht die Suche in EnArgus klar über vergleichbare Angebote hinaus, sowohl in Suchmöglichkeiten, Auswertungswerkzeugen und Intelligenz der Suchmaschine.

Ein Indikator für die Qualität der in EnArgus erstellten Ontologie zur Suche nach Projekten ist die Abdeckung der betrachteten Energieforschungsthemen durch die Begriffe der hierfür erstellten Ontologie. Dies wird überprüft, indem gezählt wird, wie viele Projekte eines vorgegebenen Teils der Leistungsplansystematik durch die Ontologie-Begriffe gefunden werden. Diese Abdeckungsquote für die in EnArgus hauptsächlich untersuchten Teile der Leistungsplansystematik liegt bei 86 % Abdeckung, d.h. 86 % der Projekte der LPS-Bereiche zur Energieforschung, die Gegenstand von EnArgus waren, werden auch gefunden. Diese Abdeckung ist

erstaunlich hoch, wenn man bedenkt, dass a) Energieforschung sehr divers ist und die geförderten Projekte auch innerhalb der Bereiche sehr vielfältig sind, b) die Leistungsplansystematik teilweise schlecht eingehalten wird, so dass Projekte in scheinbar unpassende Bereiche eingeteilt sind und c) die Abdeckungsanalyse eher restriktiv ist. Insgesamt ist der Grad der Abdeckung für die in EnArgus bearbeiteten Bereiche also als gut bis sehr gut zu bezeichnen. Im Einzelnen sind die Ergebnisse in Tabelle 2 zusammengefasst.

Vorhaben insgesamt:	*27212*
Vorhaben ausgewählt (über LPS):	3898
Vorhaben gefunden (gesamt):	17708
Vorhaben aus Zielmenge gefunden:	3351
Vorhaben nicht gefunden:	547

Tab. 2: Übersicht der Abdeckungsanalyse

Zusammenfassend lässt sich festhalten, dass EnArgus sein Ziel als ein zentrales Informationssystem für die Energieforschungsförderung erfüllt:

- Es werden bei der Suche mit EnArgus mindestens gleich viele, häufig aber mehr Ergebnisse gefunden als in einer einfachen PROFI-Suche.
- Die umfangreichen Möglichkeiten zur direkten Auswertung, Visualisierung und Export der Suchergebnisse im Recherche-Cockpit bewirken ein komfortables Arbeiten mit EnArgus und entsprechen den Anforderungen von Ministerien, Projektträgern und Energieforschern.

EnArgus kann durch seine Ontologie-basierte Suche und das Wiki sowohl von Energieforschern, als auch von Fachfremden zielführend angewendet werden.

4 Verwertung und Ausblick

Das EnArgus-Vorhaben wurde im Juni 2013 abgeschlossen und ein EnArgus-Produktionsserver wurde zum Einsatz beim Projektträger übergeben.

Bei der hier beschriebenen Erprobung und Nutzung des Systems zeigte sich bereits die Validität des Ansatzes in Bezug auf Systemantworten für die derzeit abgedeckten Bereiche der Energieforschung. Aufgrund dieser guten Ergebnisse und des bereits jetzt erkennbaren hohen Nutzens für die Anwender soll mit dem Nachfolgeprojekt EnArgus2.0, das im Juli 2013 gestartet ist, der Ausbau des Systems gewährleistet werden: dies betrifft sowohl die Themenfelder aus der Energieforschung, die in Breite und Tiefe umfassender bearbeitet werden, als auch die Weiterentwicklung des EnArgus-Gesamtsystems. Die Auswertungsergebnisse des laufenden Vorhabens dienen dabei als Anhaltspunkt für Verbesserungen und Erweiterungen am derzeitigen EnArgus-System.

EnArgus2.0 setzt sich eine Vielzahl von Innovationen zum Ziel: Ausbau des Energieforschungs-Wikis zur Erleichterung der Zusammenarbeit der beitragenden Energieforscher und zum Gebrauch in der Öffentlichkeit, ein grafischer Ontologie-Editor, über den Ergänzungen und Änderungen an der Ontologie für die Energieforscher ergonomischer angebracht werden können, automatisierte Verschlagwortung von Vorhaben-Daten, sowie technische Auswertung von Nutzerverhalten zur Trenderkennung und Systemoptimierung.

EnArgus.public wurde im April 2015 für die Öffentlichkeit freigeschaltet und ist unter http://www.enargus.de verfügbar.

Danksagung

Das Projekt EnArgus® wird vom Bundesministerium für Wirtschaft und Energie im Rahmen des Energieforschungsprogramms[14] gefördert. Förderkennzeichen 03ET1064 (EnArgus, Juli 2011 – Juni 2013) und 03ET4010 (EnArgus2.0, Juli 2013 – Dez. 2016).

Wir danken allen früheren und derzeitigen an EnArgus beteiligten Projekt- und Kooperationspartnern. Neben den Organisationen der Autoren dieses Papiers sind dies: Ruhr-Universität Bochum (Lehrstuhl Energiesysteme u. Energiewirtschaft, LEE), Forschungszentrum Jülich (Institut für Energie- und Klimaforschung – Systemforschung und Technologische Entwicklung, IEK-STE), Institut für Ressourceneffizienz und Energiestrategie (IREES), Technische Universität Darmstadt (Institut IWAR), Universität Stuttgart (Materialprüfungsanstalt, MPA), Zentrum für Beratungssysteme in der Technik, Dortmund e.V. (ZEDO), bense.com, sowie der Projektträger Jülich.

Die Projekt-Webseite ist verfügbar unter http://enargus.fit.fraunhofer.de.

Referenzen

[Ha12] Haarmann, B., Gottsmann, F. & Schade, U. (2012). How to make Ontologies self-building from Wiki-Texts. In: Arabnia, H.R., Ray, L.D. & Hashemi, R. (Eds.), *Proceedings of the 2012 International Conference on Information and Knowledge Engineering* (pp. 16-20). Las Vegas, NV: CSREA

[Wi10] Wietschel, M., Arens, M., Dötsch, C., Herkel, S., Krewitt, W., Markewitz, P., Möst, D., Scheufen, M. (2010). *Energietechnologie 2050 – Schwerpunkte für Forschung und Entwicklung.* ISI-Schriftenreihe Innovationspotenziale. Karlsruhe: Fraunhofer ISI.

[14] http://www.bmwi.de/DE/Mediathek/publikationen,did=427698.html

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmidt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Grundlagen sicherheitsgerichteter Software-Entwicklung

Katja Stecklina[1] und Clemens Passeck[2]

Abstract: Um die Entwicklung eines sicheren Produktes erfolgreich abzuschließen sind neben technischen Aspekten und Anforderungen an die Software immer auch die Vorgehensweise und die Dokumentation der Entwicklungstätigkeiten für eine Zertifizierung bzw. Zulassung sehr wichtig. Dazu gehören die systematische Dokumentation von Anforderungen, eine systematische Implementierung und umfangreiche qualitätssichernde Aktivitäten wie das Testen und die Durchführung von Reviews. Neben den klassischen Software-Entwicklungsmethoden spielen seit längerer Zeit Vorgehensweisen wie Objektorientiertes Entwickeln, Agile Software-Entwicklung und Modellbasierte Entwicklung auch in der sicherheitsgerichteten Software-Entwicklung eine Rolle. Das Tutorial bietet einen branchenübergreifenden Einblick in die Anforderungen an die Entwicklung funktional sicherer Eingebetteter Software.

Keywords: Funktionale Sicherheit in der Software Entwicklung, Prozesse zur Entwicklung von sicherheitsgerichteter Eingebetteter Software

1 Einleitung

Das Thema Funktionale Sicherheit ist inzwischen in vielen Branchen auch für Softwareentwicklungsprojekte relevant geworden. Schon lange gibt es Anforderungen an Softwareentwicklungsprozesse aus Standards wie RTCA DO-178C, IEC 61508-1 und ISO 26262. Diese Standards werden regelmäßig weiterentwickelt und genügen immer höheren Ansprüchen an die Funktionale Sicherheit. Aktuell kann man beobachten, dass viele branchenspezifische Standards veröffentlicht werden, die sehr ähnliche Anforderungen beinhalten, beispielsweise ist seit einiger Zeit Modified Condition/ Decision Coverage als Abdeckungsmaß für geprüfte Softwareteile in RTCA DO-178C, DIN EN 61508-3 und auch in ISO 26262 gefordert. Aus einem Extrakt der wesentlichen Forderungen und Ziele dieser Standards erhält man einen Überblick über notwendige Arbeiten um für eine Software eine gegebenenfalls notwendige Zulassung zu erlangen. Im Folgenden wird auf diese wesentlichen Forderungen und Ziele eingegangen.

[1] Philotech Systementwicklung und Software GmbH, Kompetenzzentrum für Funktionale Sicherheit, Bahnhofstr. 30, 03046 Cottbus, Katja.Stecklina@philotech.de
[2] Philotech Systementwicklung und Software GmbH, Fachbereich Verifikation und Validierung, Bahnhofstr. 30, 03046 Cottbus, Clemens.Passeck@philotech.de

2 Ermitteln der zu erreichenden Sicherheitsstufe

Zu Beginn der Arbeiten an einer Software sollte bekannt sein, welche Regularien und Standards für die Entwicklung herangezogen werden müssen. Im Normalfall stehen die Regularien im Vorfeld fest.

Gemäß den jeweiligen Regularien wird die Stufe der Sicherheit für die Eingebettete Software basierend auf Analysen des Gesamtsystems festgelegt. Für die Funktionen, die das System erfüllt, werden dabei die Fehlfunktionen und deren Auswirkungen untersucht. Je nach Schwere der Folgen, teilweise mit Berücksichtigung weiterer Faktoren wie Kontrollierbarkeit und Eintrittswahrscheinlichkeit von bestimmten Situationen, wird für die Funktion eine Stufe der Sicherheit bestimmt. In DIN EN 61508-1 beispielsweise werden diese „Safety Integrity Level“ genannt.

Die jeweilige Stufe der Sicherheit wird dann gemäß bestimmter Standard-abhängiger Regeln der Software und deren Funktionen zugeordnet. Auf Basis dieser Stufen werden die Maßnahmen bestimmt, die während der Software-Entwicklung ergriffen werden müssen um den Standard zu erfüllen. Man geht dann davon aus, dass eine Software, die die Vorgaben erfüllt, hinreichend sicher ist. Dieses Konzept wird im Luftfahrtbereich „Development Assurance“ genannt [RTa].

3 Der Lebenszyklus der Software-Entwicklung

Ein wesentlicher Aspekt der Development Assurance ist das Festlegen eines für das jeweilige Softwareentwicklungsprojekt passenden Lebenszyklus. Normalerweise besteht ein solcher Lebenszyklus aus mehreren Phasen, beispielsweise Anforderungsdefinition, Design, Implementierung, statische Analyse und dynamisches Testen. Für jede dieser Phasen werden Ein- und Ausgangsinformationen festgelegt, welche Aktivitäten innerhalb der Phasen stattfinden und welche Übergangskriterien zwischen den Phasen gelten müssen.

Bestimmte Anforderungen müssen in den zu definierenden Phasen des Lebenszyklus abgebildet werden. Diese Anforderungen werden in der Regel spezifisch für die Stufe der Sicherheit festgelegt und betreffen häufig die überprüfenden Aktivitäten. Beispielsweise müssen gemäß RTCA DO-178C für jede Software-Anforderung im Pflichtenheft eine adäquate Menge von Testfällen existieren, die auf der Zielhardware unter möglichst realistischen Einsatzbedingungen durchgeführt werden müssen. Während der Definition des Lebenszyklus müssen diese Anforderungen berücksichtigt werden.

Die Phasen selbst müssen mitnichten strikt in Reihenfolge oder einmalig durchgeführt werden. Iterative inkrementelle Ansätze sind gängig. In vielen Projekten werden diese

Phasen mehrfach oder parallel durchgeführt. Wichtig ist, dass die Übergangskriterien erfüllt sind bevor Aktivitäten begonnen werden und dass die Anforderungen an das Konfigurationsmanagement jederzeit erfüllt sind. Hierzu gehört zum Beispiel, dass nach der Änderung einer Anforderung je nach Fortschritt die Folgen der Änderung analysiert und dokumentiert werden müssen. Eine Anforderungsänderung kann zum Beispiel eine Teständerung oder Implementierungsänderung nach sich ziehen, oder eine Überprüfung anderer Anforderungen auf Konsistenz kann erforderlich werden.

Es ist sogar möglich, innerhalb des Projektes mehrere unterschiedliche Lebenszyklen für verschiedene Teile der Software zu definieren. Der Zukauf eines Teils der Software erfordert einen ganz anderen Lebenszyklus als eine Neuimplementierung oder die Wiederverwendung von Modulen aus anderen Projekten.

Die Definition eines oder mehrerer Lebenszyklen für ein Projekt ist nicht immer einfach. Neben den Regularien müssen auch erwartete Aufwände, Werkzeuge, Kompetenzen der Entwickler, Zeitpläne und andere Faktoren berücksichtigt werden.

4 Einbindung von speziellen Techniken und Sonderfällen

In der Praxis stellt sich heraus, dass ein Softwareentwicklungsprozess mit der gängigen Abfolge von Anforderungsdefinition, Design, Implementierung in Verbindung mit Reviews und Testen sehr gut beherrschbar ist. Nicht selten jedoch ist diese gängige Abfolge für das Projekt bzw. Teilprojekt nicht zutreffend. Oft kommt es beispielsweise vor, dass Teile der Software zugekauft oder extern entwickelt werden, Bibliotheken oder alte Softwareteile integriert werden müssen oder Werkzeuge benutzt werden sollen, die bestimmte Prozessschritte automatisieren. Oft treten in solchen Fällen Probleme auf, zum Beispiel dass Anforderungen zu wiederverwendeten Softwareteilen fehlen oder dass das Vertrauen in die automatisierenden Werkzeuge nicht vorhanden ist.

Die Behörden und Regularien bieten für solche Ausnahmen entweder direkt Lösungsansätze an, z.B. geht RTCA DO-331 auf Anforderungen an Modellbasiertes Entwickeln ein. Ansonsten ist es in der Regel ratsam, mit der zulassenden Behörde oder Institution frühzeitig das Gespräch zu suchen, wie mit den entsprechenden Forderungen in diesen Sonderfällen umzugehen ist.

5 Zusammenfassung

Der Ansatz „Development Assurance“ oder vergleichbaren Konzepten für die Produktentwicklung soll sicherstellen, dass das Verhalten der Produktsoftware mit hinreichender Wahrscheinlichkeit keine Gefahr für Menschen darstellt. Innerhalb eines geeigneten zu definierenden Lebenszyklus werden alle Anforderung an den Prozess abgebildet, die je nach Sicherheitsstufe erforderlich sind.

Literaturverzeichnis

[DIa] DIN EN 61508-1: Funktionale Sicherheit sicherheitsbezogener elektrischer/ elektronischer/programmierbarer elektronischer Systeme – Teil 1: Allgemeine Anforderungen (IEC 61508-1:2010)

[DIb] DIN EN 61508-3: Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme – Teil 3: Anforderungen an Software (IEC 61508-3:2010)

[RTa] Radio Technical Commission for Aeronautics: RTCA DO-178C: Software Considerations in Airborne Systems and Equipment Certification. Washington: RTCA, 2011 (RTCA/DO-178C)

[RTb] Radio Technical Commission for Aeronautics: RTCA DO-331: Model-Based Development and Verification Supplement to DO-178C and DO-278A. Washington: RTCA, 2011 (RTCA/DO-331)

[ISO] International Organization for Standardization: ISO 26262:2011: Road vehicles – Functional safety. Geneva, 2011

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Doktorandenprogramm der INFORMATIK 2015

Doktorandenprogramm der INFORMATIK 2015

Ansgar Scherp[1], Steffen Becker[2] und Heinrich T. Vierhaus[3]

Zusammenfassung

Das Doktorandenprogramm der INFORMATIK 2015 erlaubt jungen Forschern und Forscherinnen ihr Thema in einem Vortrag oder als Poster einem breiten Publikum vorzustellen. Es gibt den DoktorandInnen außerdem die Möglichkeit sich mit anderen Forschern und Doktoranden auszutauschen, gegenseitige Kontakte zu knüpfen, Probleme und deren Lösungswege zu diskutieren und Rückmeldungen von promovierten WissenschaftlerInnen und ProfessorInnen auf verwandten Forschungsgebieten zu erhalten. Neben den wissenschaftlichen Präsentationen und Diskussionen steht daher im Doktorandenprogramm der persönliche Kontakt und Austausch mit erfahrenen Forschern im Vordergrund.

Programmkomittee

Prof. Andreas Dengel, Technische Universität Kaiserslautern
Prof. Andreas Henrich, Universität Bamberg
Prof. Andreas Winter, Universität Oldenburg
Prof. Antonio Krüger, Saarland University
Prof. Bernd Neumann, Universität Hamburg
Prof. Bernhard Rumpe, RWTH Aachen
Prof. Christian Freksa, Universität Bremen
Prof. Collin Atkinson, Universität Mannheim
Prof. Dirk Nowotka, Universität zu Kiel
Prof. Elisabeth André, Universität Augsburg
Prof. Gregor Engels, Uni Paderborn
Prof. Gunter Saake, Universität Magedeburg
Prof. Harald Kosch, Universität Passau
Prof. Heiner Stuckenschmidt, Universität Mannheim
Prof. Ingo Schmitt, Universität Cottbus
Prof. Isabella Peters, Universität Kiel
Prof. Jürgen Ebert, Universität Koblenz-Landau

[1] ZBW – Leibniz-Informationszentrum Wirtschaft und Universität zu Kiel, asc@informatik.uni-kiel.de
[2] TU Chemnitz, steffen.becker@informatik.tu-chemnitz.de
[3] BTU Cottbus - Senftenberg, htv@informatik.tu-cottbus.de

Prof. Klaus-Dieter Althoff, Universität Hildesheim
Prof. Lars Grunske, Uni Stuttgart
Prof. Lars Schmidt-Thieme, Universität Hildesheim
Prof. Mario Schölzel, Leibniz-Institut fuer innovative Mikroelektronik, Frankfurt (Oder)
Prof. Melanie Herschel, Universität Stuttgart
Prof. Michael Gertz, Universität Heidelberg
Prof. Miriam Mehl, Universität Stuttgart
Prof. Philipp Slusallek, Universität des Saarlandes
Prof. Ralf Reussner, Universität Karlsruhe
Prof. Raúl Rojas, FU Berlin
Prof. Rolf Drechsler, Universität Bremen
Prof. Samuel Kounev, Uni Würzburg
Prof. Sebastian Rudolph, TU Dresden
Prof. Stefan Conrad, Universität Düsseldorf
Prof. Stefan Wagner, Universität Stuttgart
Prof. Stefan Wrobel, Universität Bonn
Prof. Susanne Boll, Universität Oldenburg
Prof. Sybille Hellebrand, Universitaet Paderborn
Prof. Thomas Neumann, TU München
Prof. Thomas Plagemann, Universität Oslo
Prof. Uli Sattler, University of Manchester
Prof. Viorica Sofronie-Stokkermans, Universität Koblenz-Landau
Prof. Volker Turau, TU Hamburg-Harburg
Prof. Wilhelm (Willi) Hasselbring, Universität zu Kiel
Prof. Wolfgang Effelsberg, Universität Mannheim
Prof. Wolf-Tilo Balke, Universität Braunschweig

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Effizienz im elektromobilen Massenmarkt

Malte Zuch, Hochschule Hannover, Fakultät IV – Wirtschaft und Informatik
Ricklinger Stadtweg 120, 30459 Hannover, Raum 301, Online: *www.just-charge.com*

Abstract: Elektromobilität kennzeichnet sich durch begrenzte Fahrreichweiten und lange Ladezeiten. Selbst das Laden an öffentlichen Schnellladesäulen übersteigt die reguläre Tankzeit gängiger Fahrzeuge mit Verbrennungsmotoren um ein vielfaches. Das begrenzte Angebot dieser öffentlichen Ladesäulen und die Zielsetzung der Bundesregierung, in den nächsten fünf Jahren eine Million Elektrofahrzeuge auf deutschen Straßen zu etablieren, wird die Problematik im Auffinden von freien Ladesäulen zunehmend strapazieren. Forschungsgegenstand dieser Arbeit ist ein echtzeitnahes Ladesäulen-Management-System, welches die Fahrzeuge im Auffinden von unbelegten oder demnächst frei werdenden Ladesäulen unterstützt, so dass unter anderem Fahrten zu bereits belegten Ladesäulen minimiert werden können.

Keywords: Elektromobilität, Fahrassistenzsysteme, Ladesäulenmanagement, Fahrzeugmanagement

1 Einleitung

Gängige Fahrzeuge mit Verbrennungsmotoren können in wenigen Minuten Fahrreichweiten bis zu 1000 Km in einem gut ausgebauten Tankstellennetz erwerben. Elektrofahrzeuge benötigen dem gegenüber bis zu acht Stunden Ladezeit an privaten Hausanschlüssen und erwerben dabei nur Fahrreichweiten von etwa 145 Km. Selbst das deutlich leistungsfähigere Laden an öffentlichen Ladesäulen übersteigt mit 20-60 min Ladezeit die reguläre Tankzeit gängiger Fahrzeuge mit Verbrennungsmotoren noch um ein vielfaches. Auch die Verfügbarkeit dieser öffentlichen Ladesäulen ist der Verfügbarkeit von regulären Zapfsäulen weit unterlegen. Das eingeschränkte Parkplatzangebot begrenzt besonders in verkehrslastigen Innenstädten die maximale Anzahl an installierbaren Ladesäulen. Weiterhin ist es für die Ladesäulenbetreiber oftmals finanziell vorteilhaft, ein Ladesäulenunterangebot aufrecht zu halten, um die bestehende und kostenintensiv ausgebaute Ladeinfrastruktur optimal auslasten zu können. Öffentliche Ladesäulen stellen daher tendenziell ein begrenztes Gut mit einer geringen Verfügbarkeit dar. Die Bundesregierung hat in Konkretisierung des "Nationalen Entwicklungsplanes Elektromobilität" das offizielle Regierungsprogramm bezüglich Elektromobilität bereits vorgestellt. Dabei sieht die Bundesregierung vor, bis zum Jahr 2020 eine Million Elektrofahrzeuge im deutschen Markt zu etablieren [Bdr11]. Die Nationale Plattform Elektromobilität (NPE) bestätigt als Beratungsgremium der Bundesregierung dieses Ziel auch aktuell im Fortschrittsbericht 2014 und bezeichnet die Jahre 2015-2017 als Phase des Markthochlaufes [NPE14]. Demnach wird die Fahrzeugdichte an Elektrofahrzeugen in den kommenden Jahren beständig zunehmen und damit auch die Problematik im Auffinden von unbelegten öffentlichen Ladesäulen. Fahrer von Elektrofahrzeugen unterliegen dem Informationsdefizit das sie nicht wissen, welche anderen Fahrzeuge mit ihnen aktuell im Wettbewerb um mögliche freie Ladesäulen stehen. Wenn mehrere Fahrzeuge an derselben

Ladesäule laden wollen, bekommt immer das Fahrzeug welches die Ladesäule als erstes erreicht den Ladeplatz. Sämtliche anderen Fahrzeuge müssen sich eine alternative Ladesäule suchen oder Wartezeiten in Kauf nehmen. Somit entstehen durch das Informationsdefizit der Fahrer Effizienzverluste in Form von Wartezeiten und vermeidbaren Fahrten zu bereits oder demnächst belegten Ladesäulen. Zur Minderung dieser Effizienzverluste wird aktuell an einem echtzeitnahen Ladesäulen-Management-System geforscht. Von der Hochschule Hannover unterstützen Prof. Dr.-Ing. Arne Koschel seitens der Informatik und Prof. Dr. rer. nat. Andreas Hausotter seitens der Wirtschaftsinformatik im Rahmen des kooperativen Promotionsprogramms Elektromobilität (KPE)[1] das Promotionsvorhaben vom Promovenden Malte Zuch und dessen Erforschung eines Ladesäulen-Management-Systems zur Minderung von Effizienzverluste im elektromobilen Massenmarkt.

1.1 Definition der Begrifflichkeit „Elektrofahrzeug“

Der Begriff „Elektrofahrzeug“ muss bei der Formulierung der Bundesregierung, bis zum Jahr 2020 eine Million Elektrofahrzeuge zu etablieren, relativiert werden. Dieser Begriff darf nicht mit rein batteriegetriebenen Fahrzeugen gleichgesetzt werden, wie es fachfremde Medien oft öffentlich pauschalisieren. Denn der Allgemeinbegriff „Elektrofahrzeug“ umschließt gemäß dem Nationalen Entwicklungsplan Elektromobilität sowohl rein batteriegetriebene Elektrofahrzeuge (BEV) als auch Plug-In-Hybrid-Fahrzeuge (PHEV) und Range-Extender-Fahrzeuge (REEV) [Bdr9]. Der Anteil rein batteriegetriebener Elektrofahrzeuge beträgt 45 % gemäß dem zweiten Bericht der Nationalen Plattform für Elektromobilität [NPE11]. Bezogen auf die Zielsetzung der Bundesregierung, eine Million Elektrofahrzeuge bis zum Jahr 2020 zu etablieren, beträgt der rein batteriegetriebene Anteil demnach bis zu 450.000 Fahrzeuge.

1.2 Erwartete Anzahl an Ladesäulen

Basierend auf der Richtlinie des Europäischen Parlaments und des Rates über den Aufbau der Infrastruktur für alternative Kraftstoffe gilt die allgemeine Zielsetzung, pro zehn Elektrofahrzeuge eine Ladesäule zur Verfügung zu stellen [Eur14]. Die Begrifflichkeit „Elektrofahrzeug“ bezieht sich hierbei auf die von der Bundesregierung geplanten „eine Million Elektrofahrzeuge bis 2020“, welche sowohl die rein batteriebetrieben Elektrofahrzeuge als auch die Plug-In-Hybrid-Fahrzeuge und Range-Extender-Fahrzeuge beinhaltet. Diesbezüglich ergibt sich eine Ladesäulenanzahl von 100.000 Ladesäulen im Jahr 2020.

[1] KPE: „kooperatives Promotionsprogramm Elektromobilität“
Das KPE ist ein kooperatives Promotionsprogramm mit dem Themenschwerpunkt Elektromobilität. Dabei wird das Themenspektrum Elektromobilität standortübergreifend und interdisziplinär von mehreren Promovenden untersucht.

1.3 Der „elektromobile Boom“

Gemäß dem Kraftfahrzeug-Bundesamt sind zum 01. Januar 2015 in Deutschland 18.948 rein batteriegetriebene Elektrofahrzeuge zugelassen [Kba15]. Ausgehend von diesen Bestand und dem offiziellen Ziel der Bundesregierung im Jahr 2020 eine Million Elektrofahrzeuge im deutschen Markt zu etablieren, wobei 450.000 Fahrzeuge rein batteriegetriebene Fahrzeuge darstellen, ergibt sich damit ein Ausbaufaktor von 23,74. Bezogen auf die verbleibenden fünf Jahre definiert dies ein jährliches Wachstum von 88,4 %. Dieses Wachstum rechtfertigt den Begriff *e-mobiler Boom* und ist Motivation für die Erforschung und Entwicklung von hochelastischen und skalierbaren Informationsverarbeitungssystemen für ein echtzeitnahes Ladesäulen-Management-System. Dieses Informationsverarbeitungssystem soll den volatilen Anforderungen des bevorstehenden e-mobilen Booms standhalten und Elektrofahrzeug-Fahrern in ihrer Entscheidungsfindung beim Auffinden von unbelegten oder demnächst frei werdenden Ladesäulen effektiv unterstützen und die kollektive Nachfragesituation damit optimieren.

1.4 Abschätzung der bevorstehenden Ladeaktivität

Die durchschnittliche deutsche Jahresfahrleistung betrug im Jahr 2013 pro Fahrzeug und Jahr 13.840 Km [Sta15]. Laut dem Kraftfahrzeug-Bundesamt ist der BMW i3 das im Jahr 2014 marktdominierende Elektrofahrzeug in Deutschland. Mit einer praxisnahen Reichweite von etwa 145 km je Akkuladung ergibt sich bei der durchschnittlichen Jahresfahrleistung von 13.840 Km / Jahr, dass ein Elektrofahrzeug alle 3,8 Tage neu aufgeladen werden muss. Fahrzeuge werden jedoch nicht restlos leer gefahren bis sie wieder geladen werden, sondern es wird noch mit geringfügiger Sicherheitsreichweite neu geladen. Bei 15 km verbleibender Sicherheitsreichweite ergibt sich ein Laderhythmus von spätestens 3,5 Tagen, wenn die Fahrzeuge keine Zwischenladungen vornehmen würden. Dieses Szenario aggregiert sich im Jahr 2020 auf mindestens 123.288 täglich zu ladende rein batteriebetriebene Elektrofahrzeuge, welchen 100.000 Ladesäulen zur Verfügung stehen (siehe Kapitel 1.2).

2 Abgrenzung zu anderen Arbeiten

In der Arbeit von Mariyasagayam und Kobayashi wird ein System zur Routen-Optimierung von Elektrofahrzeugen beschrieben [MK13]. Bei dieser Betrachtung werden definierte Startpunkte und Endpunkte berücksichtigt. Ausgehend von diesen Start und Endpunkten der Primär-Route werden den Elektrofahrzeugen Alternativrouten vorgeschlagen, welche in Abhängigkeit der aktuellen Verkehrslast eine potentiell bessere Ladewahrscheinlichkeit entlang der neuen Route versprechen. Spontanladen im aktuellen Verkehrshotspot (Innenstadtbereich in Großstädten) und unterschiedliche Preisstrukturen der Ladesäulenbetreiber werden nicht betrachtet, sondern die konkrete Routenplanung zum vorher bestimmten Zielort steht hierbei im Vordergrund. In einer weiteren angrenzenden Arbeit wird ein Lademanagement hinsichtlich kooperierender und egoistischen Elektrofahrzeugen bezüglich einer Preisstruktur untersucht [Ha12]. Dabei wird einigen Fahrzeugen ein kooperatives Verhalten unterstellt, welche zu fixen

Preisen Energie an einer Ladesäule nachfragen. Dem gegenüber stehen weitere Fahrzeuge mit einem egoistischen Verhalten, welche dynamisch immer den jeweils günstigsten Preis an einer Ladesäule (aus)-nutzen. Die kooperativen Fahrzeuge können weiterhin auch dem Energienetz über die Ladesäulen ihre im Fahrzeug gespeicherte Energie für eine Netzstabilisierung bereitstellen. Gegenstand dieser Betrachtung sind Simulationen, welche dabei die Profitabilität einer Ladesäule in dem Nachfrageumfeld „egoistische vs. kooperative Fahrzeuge“ optimieren.
Die allgemeine Planung einer Ladeinfrastruktur und die optimale Standortwahl für eine Ladesäule wurde in einer angrenzenden Dissertation beschrieben [Ks12]. Die optimale Standortauswahl bei einer Ladesäulenplanung stellt demnach ein NP-schweres mathematisches Problem dar, welches mit geeigneten Heuristiken mittels einer Simulation beschrieben wurde. Aufgrund des Ansatzes der Standortplanung für Ladesäulen findet eine allgemeine Betrachtung der Energienachfrage statt und keine echtzeitnahe Betrachtung.

Gegenüber den oben genannten angrenzenden Arbeiten besteht das Ziel der Arbeit welche hinter diesem Paper steht in einem echtzeitnahen Ladesäulen-Management-System für stark frequentierter Verkehrshotspots (Innenstadtbereiche), welches die Fahrer in der spontanen Ladesäulenentscheidung unterstützt. Der Fokus hierbei steht im dynamischen Vermitteln von Ladesäulen noch während der Fahrt im aktuellen Verkehrsfluss unter Berücksichtigung sämtlicher im Wettbewerb stehenden anderen Fahrzeuge.

3 Darstellung des Ladesäulen-Management-Systems

Schon heute verfügen viele Fahrzeuge über eine mobile Datenverbindung. Entweder direkt über das Bordsystem des Fahrzeugs oder indirekt über das Smartphone des Fahrers. Zukünftig kann davon ausgegangen werden, dass deutlich mehr Fahrzeuge über eine mobile Datenverbindung verfügen werden. Somit können die Fahrzeuge ihren aktuellen Standort und ihren Ladestand dem Ladesäulen-Management-System mitteilen. Das System wertet diese Daten echtzeitnah aus, um jedem Fahrzeug rechtzeitig die bestmögliche Handlungsempfehlung zum Auffinden einer aktuell freien bzw. einer demnächst frei werdenden Ladesäule mitteilen zu können. Ausgehend vom individuellen Ladestand und der Fahrzeugposition wird die höchste Ladesäulen-Erreichbarkeits-Wahrscheinlichkeit bei minimalen Fahraufwand berechnet und dem Fahrzeug mitgeteilt. Die Berechnung der Daten berücksichtigt dabei stets die Standortdaten und Ladestände sämtlicher im Wettbewerb stehenden anderen Fahrzeuge. Diese kollaborative Betrachtung ermöglicht, dass bevorstehende gegenseitige Blockierungen an Ladesäulen den Fahrzeugen rechtzeitig mitgeteilt werden können und eine effiziente Empfehlung von alternativen Ladesäulen stattfinden kann. Die Ladesäulenempfehlung wird dabei den Fahrzeugen noch dynamisch während der Fahrt im aktuellen Verkehrsfluss mitgeteilt. Hierbei ist besonders ein reaktives Antwortzeitverhalten des Systems erforderlich, damit noch rechtzeitig akzeptable Ladeempfehlungen zu den Fahrzeugen kommuniziert werden können. Wenn die Zeitspanne zwischen der Übermittlung vom aktuellen Standort und Ladestand des Fahrzeugs und der anschließenden Auswertung im Ladesäulen-Management-System zu langsam erfolgt (länger als zehn Sekunden), dann würde sich das Fahrzeug bereits an einer anderen Position befinden, an welcher es die empfohlene freie Ladesäule nicht mehr

effizient erreichen könnte. Beispielsweise könnte nach einer verzögerten Antwort von länger als zehn Sekunden das Fahrzeug bereits eine abbiegerelevante Kreuzung zu einer geeigneten Ladesäule überfahren haben, so dass sich keine effizient Wendemöglichkeit mehr für das Fahrzeug ergibt, um die Ladesäule noch rechtzeitig und mit minimalen Fahraufwand ohne Umweg zu erreichen.

3.1 Partitionierung der Problemstellung

Für das Jahr 2020 werden 123.288 täglich zu ladende rein batteriebetriebene Elektrofahrzeuge erwartet, welchen 100.000 Ladesäulen zur Verfügung stehen (siehe Kapitel 1.4). Dabei sind nicht sämtliche Fahrzeuge ausschließlich auf das Laden an öffentlichen Ladesäulen angewiesen, sondern können alternativ auch am deutlich langsameren Hausanschluss laden, sofern dieser verfügbar ist. Doch besonders in Innenstädten ist ein erreichbarer Hausanschluss aufgrund der Parksituationen oftmals nicht realisierbar. Darüber hinaus werden diese Fahrzeuge zum Teil auch zusätzlich an öffentlichen Ladesäulen laden. Somit stehen diese Fahrzeuge potentiell auch mit allen anderen Fahrzeugen im Wettbewerb um die Ladesäulen und beeinflussen die Berechnung der Ladesäulen-Erreichungs-Wahrscheinlichkeit der anderen Fahrzeuge. Daher müssen auch die Fahrzeuge mit der Lademöglichkeit am privaten Hausanschluss in die Berechnung mit einbezogen werden, sobald diese sich im Verkehrsfluss befinden. Damit die Berechnung der Ladesäulenempfehlungen unterhalb von zehn Sekunden erfolgen kann, sollte die Problemstellung partitioniert werden. Ein aufwendiges Berechnen sämtlicher möglichen Fahrzeug-Ladesäulen-Kombinationen ist aber nicht notwendig, da nur die Kombinationen berechnet werden müssen, welche innerhalb der Fahrreichweite der Fahrzeuge liegen. Praktisch reduziert sich die Problemstellung damit auf die Berechnung auf Großstadtebene. Großstädte bieten dabei das Potential, voneinander abgrenzbare Teilproblemstellungen einzeln zu berechnen, statt sämtliche Daten simultan zu berechnen. Die Verkehrssituationen können für Großstädte separat berechnet werden, da die Fahrzeuge zeitgleich nur um die lokal verfügbaren Ladesäulen innerhalb der jeweiligen Großstadt im Wettbewerb stehen, nicht aber mit den von anderen Großstädten, welche außerhalb der Fahrreichweite liegen. Somit werden nur ein Bruchteil der 123.288 Fahrzeuge und 100.000 Ladesäulen zeitgleich in Abhängigkeit zueinander berechnet. Bei aktuell 76 Großstädten müssen demnach durchschnittlich pro Großstadt nur etwa 1.600 Fahrzeuge und 1.300 Ladesäulen in Abhängigkeit zueinander berechnet werden, was die Komplexität der Berechnungen reduziert. Die empfangenen Fahrzeugdaten können dabei zunächst auf einem zentralen Server gespeichert und dann für jede Großstadt auf separaten Servern berechnet werden, wie es die nebenstehende Abbildung veranschaulicht.

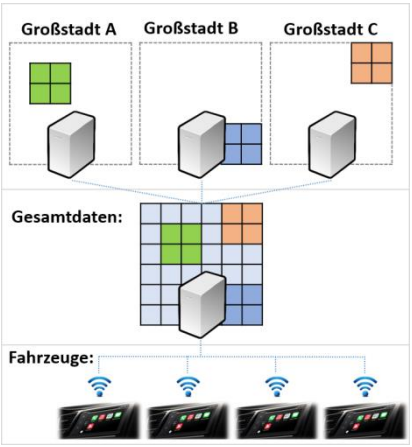


Abbildung 1: Konzept zur Partitionierung der Problemstellung

3.2 Erste Simulationen und Teilergebnisse

Erste Ergebnisse des sich noch in der Entwicklung befindenden Systems können an dieser Stelle bereits genannt werden. Simuliert wurden Großstadtbereiche mit einer Größe von 10 km x 10 km, mit 1.600 Fahrzeugen, welche im Wettbewerb um 1.300 Ladesäulen mit einer Standardladeleistung von 20 kW stehen. Simuliert wurde der „Rush-Hour"-Betrieb in dem alle Fahrzeuge in ständiger Bewegung sind und sich bei niedrigem Ladestand auf die Ladestellensuche begeben. Zunächst wurde das egoistische Standardfahrverhalten ohne aktiviertem Ladesäulen-Management-System simuliert, bei dem die Fahrzeuge immer zu der Ladesäule fahren, welche sich am nächsten zur aktuellen Fahrzeugposition befindet. Dem gegenüber wurden Simulationen mit aktiviertem Ladesäulen-Management-System getätigt. Dabei wurden unterschiedliche Adaptionsquoten der Fahrer berücksichtigt, wenn 10 %, 50 % und 100 % der Fahrer den Empfehlungen des Ladesäulen-Management-Systems adaptieren. Es wurde der durchschnittliche Ladeweg[2] der Fahrzeuge protokolliert und die durchschnittlichen Warteschlangen erfasst, welche sich an den Ladesäulen bilden. Abschließend wurde der Energiedurchsatz je Fahrzeug und Stunde erfasst.

Simulation	Ladeweg	Warteschlange	Energiedurchsatz
0 % Ladesäulen-Management-System:	2,7 %	1076	1,5 kWh/h
10 % Ladesäulen-Management-System	2,5 %	893	1,8 kWh/h
50 % Ladesäulen-Management-System	2,4 %	284	3,3 kWh/h
100 % Ladesäulen-Management-System	3,1 %	20	3,9 kWh/h

Tabelle 1: Gegenüberstellung der Simulationsergebnisse

Die Simulationen zeigen, dass ein Ladesäulen-Management-System auch schon bei geringen Adaptionsquoten der Fahrer positive Effekte erzielen kann. Wenn nur 10 % der Fahrer den optimierten Empfehlungen des Ladesäulen-Management-Systems nachkommen, können die durchschnittlich erforderlichen Ladewege für die Fahrer um -3,2 % reduziert werden. Gleichfalls nehmen die Warteschlangen an den Ladesäulen um -20,5 % ab. Der Energiedurchsatz wird um + 25,7 % verbessert.

4 Zusammenfassung und Ausblick

Der aktuelle Forschungsstand zeigt bereits erste positive Effekte, welche mit einem Ladsäulen-Management-System erzielbar sind, selbst wenn die Ladeempfehlungen des Systems nur von einer Minderheit der Fahrer befolgt werden. Im weiteren Forschungsverlauf werden die Simulationen mit weiteren Fahrprofilen ergänzt und zusätzliche Szenarien simuliert.
Darüber hinaus erfolgt eine weiterführende Untersuchung bezüglich einer effizienten Partitionierbarkeit der Gesamtproblemstellung und der zugehörigen Verteilung der zu berechnenden Daten auf unterschiedlichen Berechnungs-Servern.

[2] Ladeweg: Bei dem Ladeweg handelt es sich um den Weg, welchen die Fahrzeuge investieren müssen, um zu einer Ladesäule zu gelangen. Wenn ein Elektrofahrzeug z.B. 145 km Fahrreichweite mit einer Akkuladung hat und durchschnittlich 2 km Ladeweg zum Aufsuchen einer freien Ladesäule investieren müsste, dann würden der Ladeweg 1,37 % der Gesamtreichweite betragen. In den Simulationen wurde der Ladeweg zur besseren Interpretation daher in Relation (Prozent) zur Gesamtreichweite der Fahrzeuge angegeben.

Literaturverzeichnis

[Bdr11] Bundesregierung,
Regierungsprogramm Elektromobilität, S. 10, 2011.

[NPE14] Nationale Plattform Elektromobilität,
Fortschrittsbericht 2014 – Bilanz der Marktvorbereitung, S. 3, 2014.

[Bdr9] Bundesregierung,
Nationaler Entwicklungsplan Elektromobilität, S. 6, 2009.

[NPE11] Nationale Plattform Elektromobilität,
Zweiter Bericht der Nationalen Plattform Elektromobilität, S. 31, 2011.

[Eur14] European Union,
Directive 2014/94/EU of the european Parliament and of the Council, S. 4, 2014

[Kba15] Kraftfahrt-Bundesamt,
Bestand an Pkw am 1. Januar 2015 nach ausgewählten Kraftstoffarten, 2015.

[Sta15] Statistisches Bundesamt,
Umweltökonomische Gesamtrechnungen, S. 12, 2015.

[MK13] Electric Vehicle Route Assistance Using Forecast on Charging Station
The Third International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies, S. 139, Lisbon, 2013.

[Ha12] Han, Y.; Chen, Yan.; Han, F.; Liu, K.
An Optimal Dynamic Pricing and Schedule, University of Maryland, Maryland, S. 2, 2012.

[Ks12] Stiefen, K.
Simulation und Optimierung der Standort- und Kapazitätsauswahl in der Planung von Ladeinfrastruktur für batterieelektrische Fahrzeugflotten, Universität Paderborn, Paderborn, S. 120, 2012

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Durchsuchbare Verschlüsselung in NoSQL Datenbanken

Tim Waage[1]

Abstract: Heutzutage werden immer häufiger sog. „Cloud"-Speicherdienste eingesetzt, sei es als Backup oder um die eigenen laufenden Kosten für IT-Infrastruktur zu senken. Leider sind diese aber nicht unbedingt vertrauenswürdig, wenn es um die Behandlung sensibler Informationen geht. Eine Verschlüsselung der Daten kann dem entgegenwirken, schränkt dann aber die Möglichkeiten ein, weiterhin wie gewohnt mit der Datenbank zu interagieren. Insbesondere wird das effiziente Durchsuchen erschwert. Eine Reihe von Schemata zur durchsuchbaren Verschlüsselung nehmen sich bereits dieses Problems an, jedoch mangelt es an praktischen Implementationen und der Erprobung mit existierenden Datenbanktechnologien. Die vorliegende Arbeit präsentiert eben solche Implementationen von zwei sehr unterschiedlichen Ansätzen ([SWP00, HK14]) zur durchsuchbaren Verschlüsselung in zwei populären NoSQL Datenbanken: Apache Cassandra und Apache HBase. Dabei liegt der Fokus auf den Problemen der Schemata im praktischen Einsatz, einem Performacevergleich und der Diskussion von Optimierungsmöglichkeiten.

Keywords: Durchsuchbare Verschlüsselung, verschlüsselte Datenspeicherung, NoSQL Datenbanken, Benchmarking, Apache Cassandra, Apache HBase

1 Einleitung

Cloud Computing als Alternative zur Datenverarbeitung auf eigener Hardware befreit den Nutzer von den damit verbundenen administrativen Aufgaben und verschafft ihm zusätzliche Hardwareressourcen. Die vorliegende Arbeit betrachtet insbesondere den Anwendungsfall der Speicherung von Datensätzen in Clouddatenbanken. Da traditionelle SQL Datenbanken die Anforderungen moderner Internetapplikationen oft nicht mehr erfüllen, werden hierfür mehr und mehr sog. „NoSQL"(Not only SQL) Datenbanken eingesetzt. Diesen fehlen in der Regel aber jegliche Sicherheitsmechanismen. So stellen potenziell sowohl mutwillige Einbrecher, aber auch die Administratoren oder sogar andere Nutzer der Datenbanken eine Gefahr für vertrauliche Daten dar.

Das Hauptziel dieser Arbeit ist es daher, die Sicherheit der extern gelagerten Daten gegen nicht legitimierten Zugriff zu schützen. Dafür werden alle Daten, die den Einflussbereich des Nutzers verlassen, symmetrisch verschlüsselt. Die Datenbank enthält somit zu jedem Zeitpunkt ausschließlich verschlüsselte Daten. Die entsprechend notwendigen Schlüssel werden auf der Nutzerseite verwaltet. Es wird angenommen, dass potenzielle Angreifer die Daten zwar versuchen zu lesen, sie aber sonst nicht manipulieren (was ohnehin nur dazu führen würde, dass eine Entschlüsselung höchstwahrscheinlich nicht mehr möglich wäre, der Angreifer somit keine Kenntnis der verschlüsselten Informationen erlangt).

[1] Georg August Universität Göttingen, Gruppe: Knowledge Engineering, Goldschmidtstrasse 7, 37077 Göttingen, tim.waage@informatik.uni-goettingen.de

Die vorliegende Arbeit leistet somit folgendes:

- Implementation zweier Schemata für durchsuchbare Verschlüsselung, beide grundverschieden in ihren Ansätzen (sowohl scan-, als auch indexbasiert), inklusive der notwendigen Interaktionen zu den NoSQL Datenbanken Apache Cassandra und Apache HBase (derzeit die populärsten Datenbanken ihrer Kategorie [So])
- Quantifizierung der Performance beider Schemata in Kombination mit beiden Datenbanken und Überblick der jeweiligen Stärken und Schwächen
- Einschätzung praktischen Umsetzbarkeit, Erörterung von Optimierungspotenzialen

2 Die Datenbanken

Wie eingangs erwähnt besitzen NoSQL Datenbanken in der Regel keinerlei Sicherheitsmechanismen, wie beispielsweise eine Nutzerverwaltung und das damit verbundene Zugriffsrechtemanagement. Es wird generell angenommen, dass ein entsprechendes Frontend diese Aufgaben übernimmt. Der Einbruch in selbiges würde dann die komplette Datenbank offenlegen. Eine Möglichkeit trotzdem noch den Schutz der Daten zu gewährleisten ist die Speicherung in verschlüsselter Form. NoSQL Datenbanken bieten aber auch dafür keinerlei native Mechanismen an.

Die vorliegende Arbeit beschäftigt sich daher näher mit Apache Cassandra [LM10, Tha] und Apache HBase [Bo11, Thb]. Sie gehören sowohl zur Gruppe der Key-Value Datenbanken, als auch zu den sog. Spaltenfamiliendatenbanken. Beide wurden in Java implementiert und bieten hohe Verfügbarkeit, da sie bereits nativ ihre Daten auf vielen Servern verteilen und replizieren. Cassandra setzt dabei auf das Peer-to-Peer Prinzip, benutzt das Gossip Protokoll zur Koordination der Knoten und ist somit nach dem CAP Theorem [Br00, Br10] besonders ausgelegt für Hochverfügbarkeit und Partitionstoleranz. HBase hingegen basiert auf einer Master-Slave-Struktur. Es wird nicht unterschieden zwischen verschiedenen Datentypen wie Zahlen und Text, stattdessen wird alles als Byte-Array behandelt. Ein Zookeeper System verwaltet dabei die verschiedenen Prozesse. Bezüglich des CAP Theorems ist HBase ausgelegt für Konsistenz und Hochverfügbarkeit.

3 Die Schemata zur durchsuchbaren Verschlüsselung

Für unsere Versuche nutzen wir zwei Schemata zur durchsuchbaren Verschlüsselung. Zum einen testen wir den Algorithmus von Song et. al [SWP00] (im folgenden „SWP Schema“, in der Literatur so benannt nach den Anfangsbuchstaben der Nachnamen der Autoren). Es basiert auf linearen Scans, also dem kompletten Durchlauf der verschlüsselten Daten. Es ist eines der ersten Schemata für durchsuchbare Verschlüsselung überhaupt und auch das einzige, das in einem praktisch relevanten Kontext überhaupt schon einmal implementiert wurde [Po12]. Trotz seines auf den ersten Blick ineffizienten Linear-Scan-Ansatzes besitzt es attraktive Eigenschaften in der Praxis, die eine genauere Untersuchung durchaus rechtfertigen. Zum anderen testen wir ein relativ neues Schema, vorgeschlagen von Hahn und

Kerschbaum [HK14] (im folgenden „SUISE Schema“, kurz für „securely updating index-based searchable encryption“), welches wiederum zwei Indizes nutzt, um eine effiziente Suche zu ermöglichen.

Die folgenden beiden Kapitel stellen die Funktionsweise beider Schemata kurz vor und präsentieren einen Überblick der jeweiligen Stärken und Schwächen.

3.1 Das SWP Schema [SWP00]

Das SWP Schema ist praktisch die einzige Wahl, wenn die Aufrechterhaltung eines Index aus praktischen Gründen vermieden werden soll [Bö14]. Es untergliedert sich in vier Subschemata, beginnend mit dem „Basic Scheme“. Die *Verschlüsselung* funktioniert wie folgt: Jedes Klartextwort W wird mittels Padding oder Splitting auf eine Länge von n Bytes gebracht. Für jedes dieser Worte wird mit einem Zufallsgenerator G ein Wert S der Länge $n-m$ Bytes (mit $m < n$) erzeugt. Aus S wird dann mittels eines Schlüssels k ein (Hash-)Wert $F_k(S)$ der Länge m errechnet. $F_k(S)$ wird an S angehangen (was wieder ein Wort der Länge n ergibt) und XOR-verknüpft mit W um den finalen Geheimtext C zu erzeugen.

Die *Suche* im Schlüsseltext ist dann simpel. Wird nach einem bestimmten Wort W gesucht, so wird für jedes Wort im Geheimtext überprüft, ob $W \oplus C$ der Form $S||F_k(S)$ genügt.

Alle anderen Schemata der Autoren sind Erweiterungen dieses Grundschemas, die zusätzliche Sicherheitseigenschaften bringen sollen. Das zweite Schema („controlled searching“) erlaubt die Suche ausschließlich nach dem gewünschten Suchwort, indem es die Erzeugung von k jeweils vom Klartextwort abhängig macht. Das dritte Schema („hidden searches“) führt eine Vorverschlüsselung $E(W)$ ein, was eine Offenlegung des gewünschten Suchwortes beim Suchvorgang vermeidet. Das zweite und dritte Schema in Kombination bringen die unerwünschte Eigenschaft mit, dass die Erzeugung von k nun an vorverschlüsselte Klartextworte gebunden ist, was im Zweifel dazu führen kann, das der Dateneigentümer nicht mehr in der Lage ist, seine eigenen Daten zu entschlüsseln. Das vierte Schema („Final Scheme“) behebt dieses Problem, indem das vorverschlüsselte Wort $E(W)$ in zwei Teile L mit Länge $n-m$ und R mit Länge m gespalten wird. L kann über S rekonstruiert werden, da der Dateneigentümer den Seed des Zufallsgenerators G kennt. Aus der Errechnung von $F_k(S)$ kann dann R erschlossen werden. Aus Platzgründen sei für eine ausführlichere Darstellung an dieser Stelle auf das Originalpaper [SWP00] verwiesen.

In der vorliegenden Arbeit ist stets das „Final Scheme“gemeint, wenn auf das SWP Schema referenziert wird, da dieses alle möglichen Sicherheitsvorteile beinhaltet. Aus praktischer Sicht hat das SWP Schema folgende Stärken (+) und Schwächen (–):

\+ Aufgrund des linear-scan-Konzepts ist es nicht notwendig einen Index oder andere Statusinformationen (außer den Schlüsseln selbst) zu speichern, zu erstellen oder zu verwalten.

\+ Verschlüsselung und Suche können instantan durchgeführt werden ohne Anfragen an eine Indexstruktur stellen zu müssen.

+ Im Gegensatz zu den meisten index-basierten Schemen wird nicht nur die Information geliefert, ob ein Suchwort in einem Dokument vorhanden ist oder nicht. Auch Anzahl der Treffer und exakte Positionen im Dokument können bestimmt werden

+ Der SWP Algorithmus ist naturbedingt leicht umsetzbar für die darunterliegenden Datenbanken, bzw. generell in Client-Server-Umgebungen.

– Für große Datenmengen kann die Suche lange dauern.

– Es kann nur nach kompletten Worten gesucht werden, nicht aber nach Substrings oder regulären Ausdrücken.

– Der Geheimtext kann deutlich größer werden als der Klartext, da der SWP Algorithmus alle Klartextworte mit einer Länge von (einem Vielfachen von) n Bytes verschlüsselt (siehe Abschnitt 5.4).

3.2 Das SUISE Schema [HK14]

Im Gegensatz zum SWP Schema arbeitet SUISE mit den beiden Indizes γ_f und γ_w. γ_f speichert die verschiedenen Worte pro Dokument [2] in verschlüsselter Form. γ_w ist ein sog. invertierter Index, der zu allen bereits durchgeführten Suchen die Identifikatoren (im Folgenden kurz „IDs“) der jeweils gefunden Dokumente speichert, was wiederholtes Suchen nach dem gleichen Wort extrem beschleunigt.

Während der *Verschlüsselung* wird eine Liste aller verschiedenen Worte $(w_1, ..., w_n)$ pro Dokument f kreiert. Ein Zufallsgenerator G erzeugt die gleiche Anzahl an Werten $(s_1, ..., s_n)$. Dann werden für jedes Wort w die folgenden Schritte ausgeführt: (1) mittels einer Zufallsfunktion F und eines Schlüssels k_1 wird ein Such-Token $r_w = F_{k_1}(w)$ erzeugt. (2) Wurde r_w schon einmal benutzt, wird es zur Liste x hinzugefügt. (3) Mittels eines Zufallsorakels H wird $c = H_{r_w}(s)||(s)$ erzeugt. Die Datenbank speichert dann alle c's des Dokuments f in γ_f. Zusätzlich wird die ID des Dokuments $ID(f)$ in γ_w abgelegt für alle Einträge in x

Für die *Suche* nach Dokumenten, die ein Wort w beinhalten, wird dann zuerst wieder das Such-Token $r_w = F_{k_1}(w)$ berechnet. Existiert bereits ein Eintrag für r_w in γ_w (d.h. es wurde schon einmal nach w gesucht), kann einfach die in γ_w gespeicherte Liste der entsprechenden IDs geliefert werden. Andernfalls wird diese Liste wie folgt erzeugt: jedes c in γ_f wird zerlegt in $l||r$. Falls $H_{r_w}(r) = l$ wird die entsprechende Dokument-ID zur Ergebnismenge aufgenommen und der entsprechende Eintrag in γ_w erzeugt. Für eine ausführlichere Darstellung sei erneut auf das Originalpaper verwiesen [HK14].

Wie für das SWP Schema folgt eine Auflistung der Stärken (+) und Schwächen (–):

+ Wiederholtes Suchen nach einem Wort w ist in konstanter Zeit möglich.

+ Die Verschlüsselung benötigt nur so viele Iterationen, wie es verschiedene Worte pro Dokument gibt. Im Gegensatz dazu muss SWP alle Verschlüsselungsschritte für alle Worte ausführen.

[2] gleiche Worte im selben Dokument werden also nur als ein Eintrag im Index abgelegt

– Trotz Indizes erfordert die Suche nach noch nicht zuvor gesuchten Worten linearen Aufwand.

– Der Index γ_f speichert $c = H_{r_w}(s)||(s)$ für alle verschiedenen Worte pro Dokument. c ist somit fast immer deutlich länger als das ursprüngliche Klartextwort, wodurch γ_f sehr viel Speicherplatz belegen kann (siehe Abschnitt 5.5).

– Es werden nur die verschiedenen Worte pro Dokument erfasst, d.h. SUISE kann nicht wie SWP die genaue Anzahl und Positionen des Suchworts im Dokument liefern.

– Auch SUISE beherrscht keine Substrings oder regulären Ausdrücke.

4 Implementation

Beide Schemata wurden in Java Version 8 implementiert. Die Java Cryptography API (JCA) ist die Basisarchitektur für kryptographische Algorithmen in Java. Sie spezifiziert Interfaces, die dann wiederum von sog Kryptographie Providern implementiert werden. In der vorliegenden Arbeit kommen zwei davon zum Einsatz: die Java Cryptography Extension (JCE, die built-in Lösung von Java), und das „Legion of the Bouncy Castle package“(BC) [Thc]. Die BC-Implementationen sind fast immer etwas schneller als die der JCE, jedoch in ihrer Funktionalität meist nicht ganz so umfangreich.

Falls beide Pakete die gewünschte Funktionalität zur Verfügung stellten, wurde für die vorliegende Arbeit nach einigen Tests immer die schnellere beibehalten. Die Tabellen 1 und 2 bieten einen Überblick über die jeweils benutzen Implementierungen.

Tab. 1: Kryptographische Verfahren im SWP Schema

Funktionalität	Provider
Suchworte vorverschlüsseln $E(W)$	Bouncy Castle (AES)
Schlüssel für „controlled searching“ berechnen	JCE (HMAC-MD5)
Padding der Klartextworte auf Länge n	Bouncy Castle (PKCS7 Padding)
Zufallsgenerator G	JCE (SecureRandom, SHA1PRNG)
(Hash-)Werte $F(S)$ berechnen	JCE (HMAC-MD5)

Tab. 2: Kryptographische Verfahren im SUISE Schema

Funktionalität	Provider
Dateien verschlüsseln	Bouncy Castle (AES)
Such-Token r_w erzeugen	JCE (HMAC-SHA1)
Zufallsorakel H	JCE (HMAC-SHA1)
Zufallsgenerator G	JCE (SecureRandom, SHA1PRNG)

Zur Anbindung an Cassandra kommt der Java Treiber 2.1 [Da] zum Einsatz, der die Nutzung der aktuellen Cassandra Query Language (CQL3) erlaubt. HBase bietet als schnellste Möglichkeit der Interaktion eine native Java API, die aber ohne die Highlevel-Optionen

einer Anfragesprache wie CQL auskommen muss. Hier werden alle Anfragen auf die drei Standardoperationen put, get und delete zurückgeführt, die mit sog. Filtern dann direkt an die Tabellen und Keyspaces repräsentierenden Objekte übergeben werden.

5 Benchmarks

Das folgende Kapitel quantifiziert die Performance beider Schemata in Kombination mit beiden Datenbanken. Darüber hinaus werden alle Tests auch ohne Datenbanken auf dem bloßen Dateisystem ausgeführt, um zu ermitteln, welchen Geschwindigkeitsvorteil der Einsatz einer Datenbank überhaupt bringt. Hardwarebasis ist ein Intel Core i7-4600 CPU (2,1 GHz), 8 GB RAM und einer Samsung PM851 SSD. Darauf läuft Ubuntu 14.04 LTS, Java 8 (64bit), Apache Cassandra in Version 2.1.2, der Cassandra Java Treiber in Version 2.1, CQL 3.1.7, sowie HBase in Version 0.94.20.

5.1 Der Datensatz

Der Testdatensatz ist eine Teilmenge des TREC 2005 Spam Track Public Corpus [CL05], einer Sammlung von annähernd 100.000 E-Mails, die in ihrer jeweiligen Größe von vorwiegend einigen Kilobytes bis hin zu mehreren Megabytes variieren. Um den Anwendungsfall einer Mailbox zu simulieren, beginnen wir unsere Tests jeweils mit den ersten 1.000 Mails des Corpus und erhöhen schrittweise bis auf 10.000 Mails. Dabei wächst die Größe des Klartextes von ca. 700.000 auf 7 Millionen Worte an. Zählt man gleiche Worte pro Dokument nur einmal, beträgt die Menge der Klartextworte jeweils nur immer ca. 40%, konkret 273.204 Worte bei 1.000 E-Mails und knapp 3 Millionen Worte bei 10.000 E-Mails. Dies ist wichtig für die Einschätzung der Ergebnisse des SUISE Schemas, da dieses nur eben diese 40% des Datensatzes berücksichtigen muss, während das SWP Schema eine komplette Verschlüsselung des Datensatzes vornimmt.

5.2 Test 1: Verschlüsseln

Der erste Test misst die zur Verschlüsselung benötigte Zeit, inklusive Speicherung der Daten in der jeweiligen Datenbank oder dem Dateisystem. Diese bestehen bei SWP nur aus den verschlüsselten Dateien selbst, während bei SUISE noch die Indizes γ_f und γ_w hinzukommen. Für SWP wird die Wortlänge auf $n = 8$ festgelegt (siehe Abschnitt 5.4).

Abbildung 1 zeigt die Ergebnisse. Da beide Algorithmen einmal über den kompletten Testdatensatz iterieren müssen, wächst der zeitliche Aufwand dafür linear. SUISE ist dabei unabhängig von der darunterliegenden Datenbank schneller als SWP, obwohl komplexere Schritte zur Verschlüsselung erforderlich sind. Wie oben erwähnt ist die Ursache hierfür in der Charakteristik des Testdatensatzes zu suchen, denn wie oben beschrieben muss SUISE nur 40% der Klartextworte tatsächlich verschlüsseln.

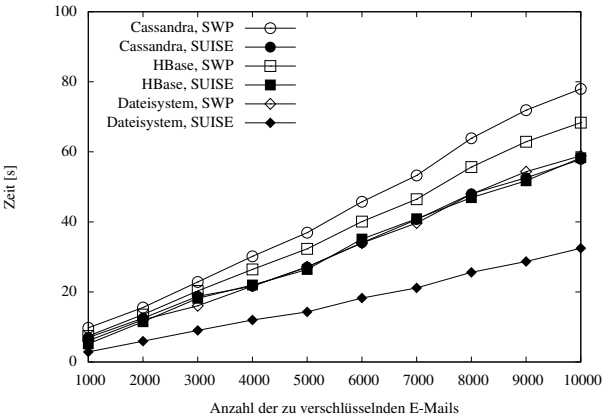


Abb. 1: Zur Verschlüsselung benötigte Zeit mit zunehmender Größe des Datensatzes

Beide Algorithmen sind schneller, wenn sie nur auf dem Dateisystem operieren, statt ihre Ergebnisse in einer Datenbank unterzubringen. Ursächlich kann angenommen werden, dass direktes Schreiben in Dateien zügiger ist, als im Falle von Cassandra erst entsprechende Insert/Update-Queries zu erstellen und auszuführen bzw. im Falle von HBase erst die entsprechenden Tabellen- und Filterobjekte zu erzeugen. Das Rennen unter den Datenbanken entscheidet HBase für sich, wenn auch nicht signifikant.

5.3 Test 2: Suchen

Der zweite Test misst die Zeit für den Suchprozess. Für einen fairen Vergleich beider Schemata wird hier aber der SWP Algorithmus insofern modifiziert, als dass er die Suche nach dem ersten Treffer in einem Dokument abbrechen darf. Damit liefert er die gleiche Information wie SUISE (nämlich ob ein Suchwort in einem Dokument vorkommt, oder nicht). Die Fähigkeit des SWP Algorithmus im Gegensatz zu SUISE eigentlich auch Anzahl und Position von Treffern pro Dokument zu liefern, würde ihm sonst zum Nachteil werden.

Abbildung 2 zeigt die Ergebnisse. Ausdrücklich angemerkt sei hier, dass im Falle von SUISE zum ersten mal nach dem Suchwort gesucht wird, also noch kein entsprechender Eintrag in γ_w vorliegt, mit dem das Suchergebnis sofort zurückgegeben werden könnte. Das resultiert ähnlich zu SWP in linear wachsender Zeit mit wachsender Datensatzgröße. Darüber hinaus macht es im Falle von SUISE in Kombination mit der Operation auf dem Dateisystem einen signifikanten Unterschied, ob die Indizes γ_f und γ_w bereits im Hauptspeicher sind (SUISE(2)), oder nicht (SUISE(1)). SUISE(2) kann somit auch als Simulation eines konstant laufenden Servers betrachtet werden, welcher die Indizes permanent im Hauptspeicher hält.

Im Gegensatz zu den Resultaten aus Test 1 ist SUISE nicht in jedem Fall besser. Der wie oben beschrieben modifizierte SWP-Algorithmus findet die gesuchten Dokumente schneller, es sei denn Cassandra wird als darunterliegende Technologie eingesetzt. Eine

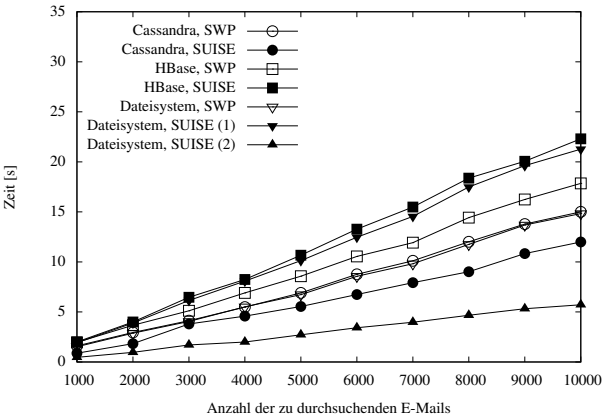


Abb. 2: Zur Suche benötigte Zeit mit zunehmender Größe des Datensatzes

Ausnahme ist die wiederholte Suche nach dem gleichen Suchwort. SUISE kann hier auf seinen Index γ_w zurückgreifen und somit in konstanter Zeit Ergebnisse liefern. Die hierfür benötigten Zeitspannen belaufen sich im Durchschnitt und unabhängig von der Datensatzgröße auf 22ms mit Cassandra und 10ms mit HBase. Bei der Operation auf dem Dateisystem und mit den Indizes im Hauptspeicher (also SUISE (2) entsprechend) ist sogar nur <1 ms notwendig. Muss erst der Index in den Hauptspeicher geladen werden (vgl. SUISE(1)) ist keine konstante Suchzeit mehr möglich. Die Suchdauer steigt dann von 1,56s für 1.000 Mails auf 17,32s für 10.000 Mails.

Was den Blick auf die Datenbanken anbelangt, so ist im Gegensatz zu den Ergebnissen beim Verschlüsselungsprozess Cassandra schneller, in Kombination mit SWP um ca. 10% und mit SUISE sogar fast doppelt so schnell verglichen mit HBase.

5.4 Test 3: Optimierung des SWP Schemas

Sowohl SWP als auch SUISE bieten Optimierungspotenzial. Bei SWP ist der wichtigste Parameter die eingesetzte Wortlänge n. Worte mit einer Länge kleiner n müssen via Padding künstlich vergrößert werden. Worte mit einer Länge größer als n müssen (mehrfach) gesplittet werden. Damit führt Padding zu mehr Bytes und Splitting zur mehr Worten im Geheimtext (und damit zu mehr benötigten Iterationen des Verschlüsselungs- und Suchalgorithmus). Das bedeutet in der Praxis: kleine n führen zu Performanceverlust, aber sparen Speicherplatz. Große n resultieren in mehr Speicherplatzverbrauch, aber können Verschlüsselung und Suche beschleunigen, da weniger Iterationen notwendig sind.

Abbildung 3 zeigt den Einfluss der wachsenden Wortlänge von $n = 4$ bis $n = 9$. Während der Speicherplatzbedarf wie erwartet mit wachsendem n linear zunimmt, stellt sich ab $n = 8$ keine signifikante Verbesserung der Performance mehr ein. $n = 8$ scheint somit der beste Kompromiss zu sein und damit geeignet als Parameter für die vorangegangenen Tests. Die optimale Wahl von n ist natürlich individuell vom konkreten Datensatz abhängig.

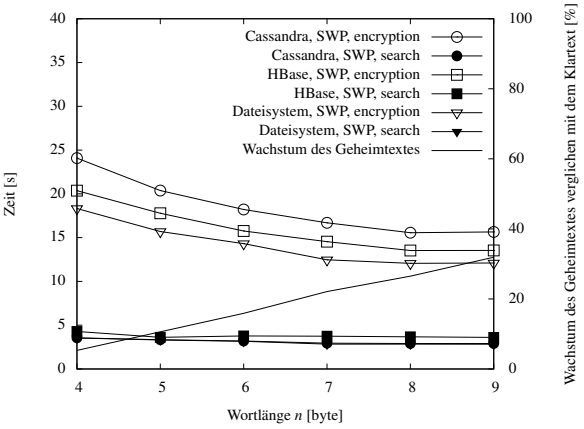


Abb. 3: Performance und Speicherplatzbedarf von SWP mit steigender Wortlänge

5.5 Test 4: Optimierung des SUISE Schemas

Die vorangegangenen Tests demonstrierten weitestgehend den Geschwindigkeitsvorteil des SUISE Schemas, der jedoch den Preis der Aufrechterhaltung zweier potenziell großer Indizes hat. Während über γ_w schwer allgemeine Aussagen getroffen werden können, da dessen Größe maßgeblich vom Suchmuster abhängt, wächst γ_f im Vergleich zum Klartext sehr schnell. Die Ursache dafür liegt in der Art und Weise, wie die Repräsentationen der Klartextworte zustande kommen, die in γ_f gespeichert werden, wie oben beschrieben nämlich $H_{r_w}(s)||(s)$. Wird wie von den Autoren vorgeschlagen [HK14] zur Umsetzung von H HMAC-SHA1 benutzt, schlägt der erste Teil bereits mit 20 Bytes zu Buche. Dazu kommt die Outputlänge des Zufallsgenerators G, der s generiert. Der Empfehlung der Autoren folgend fallen auch dafür 20 Bytes an. Das bedeutet in der Praxis: selbst wenn ein Klartextwort nur ein Byte lang ist, belegt es 40 Bytes in γ_f. Somit wird γ_f schnell sehr groß, obwohl gleiche Worte immer nur einmal pro Dokument dort abgelegt werden.

Abbildung 4 zeigt die tatsächliche Größe von γ_f abhängig von der darunterliegenden Datenbank oder dem Dateisystem. Die Größe des Testdatensatzes beträgt dabei konstant 2.000 E-Mails. γ_f wächst erwartungsgemäß linear mit der Outputlänge von G (also der Länge von s). Den empfohlenen Werten der Autoren folgend wird γ_f schon mehr als doppelt so groß als der komplette Klartext. In der Praxis kommen dazu noch die verschlüsselten Daten selbst. Verheerend ist die Situation, wenn γ_f in HBase abgelegt wird, wo die interne Ablage des Index besonders viel Platz benötigt, teilweise bis zum 5-fachen der Klartextgröße. Eine Verringerung der Outputlänge von G verringert den Platzbedarf, aber nicht signifikant. Ohne dafür gesondert eine Abbildung zu bemühen, sei des Weiteren angemerkt, dass verschiedene Outputlängen von G keinen Einfluss auf die Geschwindigkeit des SUISE Schemas haben.

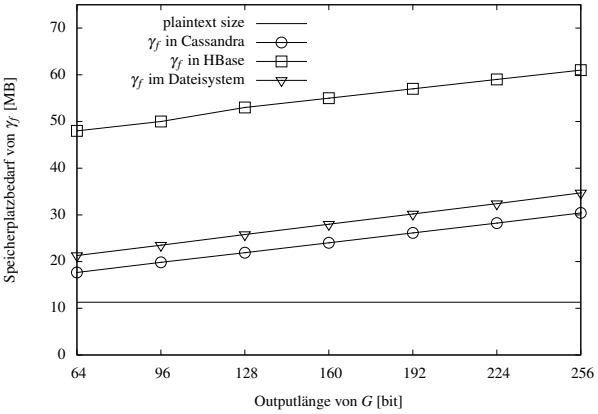


Abb. 4: Größe von γ_f mit wachsender Länge von s

5.6 Zusammenfassung

Die vorangegangenen Tests zeigen Stärken und Schwächen, sowohl der Schemata für durchsuchbare Verschlüsselung, als auch der darunterliegenden NoSQL-Datenbanken selbst. Während der *Verschlüsselung* ist der SUISE Algorithmus schneller, um den Faktor 1,2 bis 1,4 auf den Datenbanken, um den Faktor 1,8 auf dem Dateisystem. Ein ähnlich klares Ergebnis ist beim *Suchen* nicht festzustellen. Eine Ausnahme sind Anwendungsfälle, in denen mit hoher Wahrscheinlichkeit mehrfach nach denselben Worten gesucht wird. Hier ist SUISE dank seines Index γ_W zu bevorzugen.

Ansonsten kann in Anbetracht der Ergebnisse kaum eine allgemein gültige Empfehlung ausgesprochen werden. SUISE ist fast immer schneller, benötigt aber potenziell sehr große Indizes. Ist von kleineren Datensätzen auszugehen und der Speicherplatz eher knapp bemessen, so liefert auch das SWP Schema im Vergleich eine gute Performance. Dieses hat außerdem den Vorteil, dass es aufgrund seines simplen Designs einfach in Client-Server-Anwendungen umzusetzen ist.

Aus Sicht der Datenbanken ist festzuhalten, dass HBase schneller ist beim Verschlüsseln und Cassandra beim Suchen, letzteres besonders wenn der SUISE Algorithmus zum Einsatz kommt. Eine Ausnahme bildet die wiederholte Suche nach Worten mit SUISE, wo sich ein umgekehrtes Bild ergibt. Es kann außerdem beobachtet werden, dass sich zumindest für den Prozess der Verschlüsselung kein Geschwindigkeitsvorteil aus der Nutzung eines der beiden Datenbanksysteme ergibt.

6 Verwandte Arbeiten

Zu den praktischen Aspekten durchsuchbarer Verschlüsselung gibt es kaum verwandte Arbeiten. Eine Ausnahme ist [Po12], in der das SWP Schema für relationale Datenban-

ken aufgegriffen wird. Dabei gäbe es noch einige andere Schemata zu untersuchen. Eine exzellente Übersicht bietet [Bö14].

Bereits die Autoren des SWP Schemas schlagen die Nutzung eines Index für ihr Schema vor, gehen aber nicht im Detail darauf ein. Viele Nachfolgearbeiten basierten dann auf einem invertierten Index, beispielsweise [Cu06, CK10, Ca13]), da dieser auch noch effizient sein kann, wenn er nicht mehr in den Hauptspeicher passt [Ca14] (vgl. Abschnitt 5.5). Sind die zu speichernden Daten sehr dynamisch, ist dies eine besondere Herausforderung für den Index [KPR12, Ca14], während sequential-scan-Verfahren wie der SWP-Algorithmus damit naturbedingt keinerlei Probleme haben.

7 Einbettung in das übergeordnete Thema der Dissertation

Die vorliegende Arbeit zeigt, dass es mit jeweils vollkommen unterschiedlichen Ansätzen möglich ist, Suchvorgänge in verschlüsselten NoSQL-Datenbanken effizient durchzuführen. Beide Schemata erreichen bereits auf gewöhnlicher Notebookhardware ca. 100.000 Worte pro Sekunde beim Verschlüsseln und über 500.000 Worte pro Sekunde beim Durchsuchen.

Ziel der Dissertation ist es jedoch, nicht nur in verschlüsselten Daten in NoSQL-Datenbanken suchen zu können, sondern in Anlehnung an [Po12] auch andere Funktionalitäten auf ausschließlich verschlüsselten Daten bereitzustellen, welche die Query-Mechanismen dieser Datenbanken naturgemäß zu Verfügung stellen. Dazu gehört beispielsweise die Nutzung von deterministischer Verschlüsselung für Gleichheitsabfragen, Gruppierungen etc., sowie ordnungsbewahrender Verschlüsselung für sog. Range Scans, Sortierungen, dem Finden von Minima und Maxima etc. Letztere ist insbesondere für Rowkeys (zur Gewährleistung der Lokalität der Daten) und Timestamps (zur Versionierung) von Bedeutung.

Die größte Herausforderung dabei ist es, die Datenbanken an sich unverändert zu belassen. Interna wie Kompaktierungen, Bloomfilter, sog. Tombstones (Löschmarkierungen), sekundäre Indizes und andere Mechanismen sollen ungeachtet des verschlüssselten Inhalts weiterhin uneingeschränkt funktionieren.

Diese Arbeit wurde von der DFG gefördert. Förderkennzeichen: WI 4086/2-1.

Literaturverzeichnis

[Bo11] Borthakur, Dhruba; Gray, Jonathan; Sarma, Joydeep Sen; Muthukkaruppan, Kannan; Spiegelberg, Nicolas; Kuang, Hairong; Ranganathan, Karthik; Molkov, Dmytro; Menon, Aravind; Rash, Samuel et al.: Apache Hadoop goes realtime at Facebook. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, S. 1071–1080, 2011.

[Bö14] Bösch, Christoph; Hartel, Pieter; Jonker, Willem; Peter, Andreas: A survey of provably secure searchable encryption. ACM Computing Surveys (CSUR), 47(2):18, 2014.

[Br00] Brewer, Eric A: Towards robust distributed systems. In: PODC. S. 7, 2000.

[Br10] Brewer, Eric: A certain freedom: thoughts on the CAP theorem. In: Proceedings of the 29th ACM SIGACT-SIGOPS symposium on Principles of distributed computing. ACM, S. 335–335, 2010.

[Ca13] Cash, David; Jarecki, Stanislaw; Jutla, Charanjit; Krawczyk, Hugo; Roşu, Marcel-Cătălin; Steiner, Michael: Highly-scalable searchable symmetric encryption with support for boolean queries. In: Advances in Cryptology–CRYPTO 2013, S. 353–373. Springer, 2013.

[Ca14] Cash, David; Jaeger, Joseph; Jarecki, Stanislaw; Jutla, Charanjit; Krawczyk, Hugo; Rosu, Marcel-Catalin; Steiner, Michael: Dynamic searchable encryption in very large databases: Data structures and implementation. In: Proceedings of the Network and Distributed System Security Symposium (NDSS). Jgg. 14, 2014.

[CK10] Chase, Melissa; Kamara, Seny: Structured encryption and controlled disclosure. In: Advances in Cryptology-ASIACRYPT 2010, S. 577–594. Springer, 2010.

[CL05] Cormack, Gordon V; Lynam, Thomas R: TREC 2005 Spam Track Overview. In: TREC. 2005.

[Cu06] Curtmola, Reza; Garay, Juan; Kamara, Seny; Ostrovsky, Rafail: Searchable symmetric encryption: improved definitions and efficient constructions. In: Proceedings of the 13th ACM conference on Computer and communications security. ACM, S. 79–88, 2006.

[Da] Datastax: Java Driver 2.1 for Apache Cassandra. `http://www.datastax.com/documentation/developer/java-driver/2.1/common/drivers/introduction/introArchOverview.html`, Eingesehen am 04.05.2015.

[HK14] Hahn, Florian; Kerschbaum, Florian: Searchable Encryption with Secure and Efficient Updates. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. ACM, S. 310–320, 2014.

[KPR12] Kamara, Seny; Papamanthou, Charalampos; Roeder, Tom: Dynamic searchable symmetric encryption. In: Proceedings of the 2012 ACM conference on Computer and communications security. ACM, S. 965–976, 2012.

[LM10] Lakshman, Avinash; Malik, Prashant: Cassandra: a decentralized structured storage system. ACM SIGOPS Operating Systems Review, 44(2):35–40, 2010.

[Po12] Popa, Raluca Ada; Redfield, Catherine; Zeldovich, Nickolai; Balakrishnan, Hari: CryptDB: Processing queries on an encrypted database. Communications of the ACM, 55(9):103–111, 2012.

[So] SolidIT: DB-Engines Ranking. `http://db-engines.com/en/ranking`, Eingesehen am 04.05.2015.

[SWP00] Song, Dawn Xiaoding; Wagner, David; Perrig, Adrian: Practical techniques for searches on encrypted data. In: Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on. IEEE, S. 44–55, 2000.

[Tha] The Apache Software Foundation: Apache Cassandra. `http://cassandra.apache.org`, Eingesehen am 04.05.2015.

[Thb] The Apache Software Foundation: Apache HBase. `http://hbase.apache.org`, Eingesehen am 04.05.2015.

[Thc] The Legion of the Bouncy Castle. `http://bouncycastle.org/`, Eingesehen am 04.05.2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Hardware supported efficient accelerator partitioning for workstation consolidation and virtualization

Johannes Hiltscher[1]

Abstract: Accelerators have gained an important role in recent years. While being used primarily in the scientific community in the beginning, they are now employed in a wide range of every day applications. Accelerators can hence be viewed in the focus of machine consolidation and virtualization, offering new opportunities for cost saving and services. Although these opportunities have been discussed intensively, today's commercially available accelerators are not well suited for such scenarios. More specifically, they lack support for partitioning, i.e. the concurrent assignment of subgroups of their computation elements to different applications. Mechanisms for indirect partitioning support exist but have inherent limitations. This paper presents a hardware supported approach for efficient accelerator partitioning which overcomes the limitations of existing solutions. SR-IOV is used to create an accelerator able to enforce guarantees on parameters like memory bandwidth.

Keywords: Accelerator, high performance computing, IaaS, PCI Express, SR-IOV, virtualization

1 Introduction

Accelerators became widely known in recent years since graphics hardware manufacturer NVIDIA enabled programmers to run arbitrary computations on their Graphics Processing Units (GPUs). Their so called Compute Unified Device Architecture (CUDA) lets users utilize the high numerical performance of GPUs not only for rendering graphics but for arbitrary computations. As a commodity product, GPUs are considerably cheaper than other accelerator co-processors available when CUDA was released. Due to their excellent price to performance ratio, GPU based accelerators were quickly adopted in the scientific community. Today even dedicated accelerator boards are available, the NVIDIA Tesla series and the Intel Xeon Phi, the latter of which has explicitly been developed as an accelerator and is marketed as such. Rather than commodity GPUs these devices are not primarily built for graphics rendering and are thus built without graphics connectors.

Besides the scientific community, accelerators are now being used in many every day and industrial scenarios. Examples are computer aided design and construction [CK09], complex numerical simulations [Ha04], de- and encryption of data [Ma07], medical imaging [Lu08] or de- and encoding of multimedia data. To avoid the costs of having an accelerator installed in every workstation, virtualization has become increasingly popular in recent years. Using virtualized hardware it is possible to consolidate several computers into one, often more powerful, machine. This approach can help cut costs on both equipment and power. Furthermore it gives rise to new business models such as Infrastructure

[1] Technische Universität Chemnitz, Fakultät für Informatik, Professur Rechnerarchitektur und Mikroprogrammierung, Straße der Nationen 62, 09111 Chemnitz, johannes.hiltscher@informatik.tu-chemnitz.de

as a Service (IaaS, a definition can be found e.g. in [Va08]) where customers rent virtual computers on demand. As stated above, several Virtual Machines (VMs) are run concurrently on one physical machine. Each of the VMs holds a share of the physical machine's resources like processor cores and memory. The physical machine's resources are said to be partitioned in the scope of this paper.

Resource partitioning works quite well for processor cores and memory, thanks to hardware support added by processor manufacturers over recent years. In contrast, there is no such hardware virtualization support in the available accelerators, which complicates partitioning their resources. Due to the lacking virtualization support it is almost impossible to enforce Service Level Agreements (SLAs) for partitions. Furthermore accelerators even expose severe security issues [BKS13, Ma14] in multi tenant environments. In recent years there have been efforts to add partitioning support for accelerators, all of which are pure software solutions with some inherent disadvantages. An overview of prior work on accelerator partitioning and other related topics is given in Section 2. Shortcomings of the presented approaches are also analyzed in that section.

Afterwards a hardware based approach is presented in Section 3 and mechanisms to address the previously carved out problems are discussed. The presented approach is based on the PCI Single Root I/O Virtualization (SR-IOV) concept. As this paper presents a concept which is work in progress, the methodology of conceptualization, realization and evaluation is discussed in Section 4. Finally the presented concept is integrated into the context of the dissertation in Section 5.

2 State of the art and related research

Virtualization of accelerators has sparked some research and development efforts in recent years. Many of those efforts have been spent on GPU virtualization. Solutions found here are usually equally viable for accelerators due to their close relationship. One mechanism to implement device partitioning for VMs, presented in [DS09], is to create virtual devices in the hypervisor. Each VM can be assigned such a virtual device which is controlled by a special guest driver. The guest driver intercepts all critical operations and is referred to as a front-end which communicates with a back-end in the hypervisor. Intercepted operations are delegated to the hypervisor back-end which either forwards them to the actual device driver or emulates the required functionality. As functionality of the device driver is delegated to the hypervisor, this mechanism is referred to as driver delegation in this paper. A generalized depiction of this process can be found in Figure 1. Several virtual devices can be mapped to one physical device using time slicing. This approach has also been used for GPU based accelerators, as shown in [Go13, Mo12].

A similar concept was chosen in [Du10], but instead of employing a front-end driver, a reimplementation of the CUDA Application Programming Interface (API) acts as the front-end. Any calls to API functions are intercepted by the front-end middleware and forwarded to a back-end running on the machine which hosts the physical accelerator device. There the actual API function is called using the parameters received from the front-end.

Data returned from the function call is then transmitted back to the front-end. In general this mechanism - referred to as API delegation here - is very similar to driver delegation and can be visualized as the generalized delegation process shown in Figure 1. Although

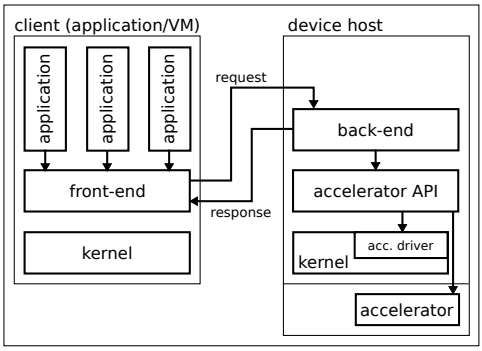


Figure 1: Generalized mechanism of API or driver delegation.

applicable to accelerator partitioning, delegating approaches inherit some drawbacks due to their realization. The added software layer (front-end and back-end) and the additional data exchange alter the latency of interactions with the device. The presented approaches are also susceptible to changes in the API or driver which may require re-engineering of the middleware. The availability of required information and a good documentation of interfaces is of paramount importance to the viability of such approaches. Further problems arise from the fact that time sharing is used to implement partitioning. This leads to the same conflicts as time sharing on general purpose processors, like cache interferences (refer to e.g. [TFJ94] for a definition) and resource conflicts. These effects can degrade performance considerably when an accelerator is shared between several applications [GWO14] which renders SLAs impossible.

Besides software solutions, hardware mechanisms are available to implement partitioning. The PCI Special Interest Group (PCI-SIG) has developed concepts on how to extend the PCI Express (PCIe) standard to allow for direct access to devices from VMs. Two variations exist for systems where multiple root complexes are present (multi root, MR) and such where only one root complex exists (single root, SR). This so-called Input/Output Virtualization - the variations of the specification are hence referred to as MR- or SR-IOV - is an extension of the PCIe specification [PC10]. As most systems are SR, the relevant document is the SR-IOV specification [Si10]. This specification defines how a device can implement so-called Virtual Functions (VFs) which are copies of the Physical Functions (PFs) of a PCIe device. For each VF some of the device hardware may be replicated, while other parts of the device's hardware are shared. An example can be a network adapter which has dedicated send and receive queues for each VF but only one shared network port. The general concept of the SR-IOV mechanism is illustrated in Figure 2. It can be seen that VM guest operating systems can directly access VFs, the logical structure of which resembles that of the respective PF they virtualize. Therefore the same device driver can be used for both hypervisor and VM guest systems. This driver only requires some added functionality for the hypervisor instance which manages the PFs of the device

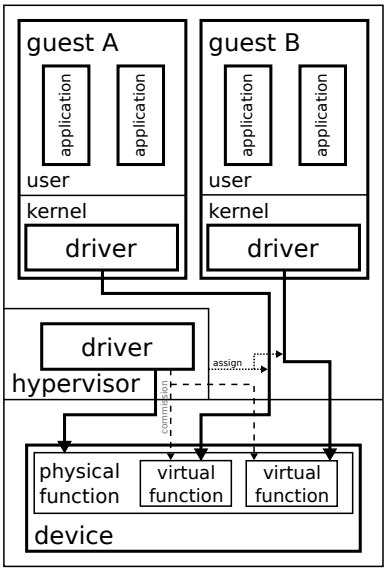


Figure 2: Concept of PCIe IOV.

and commissions the VFs. The direct access to VFs from a VM guest is possible as all critical hardware components are duplicated in the device. The hypervisor is hence not required to synchronize concurrent accesses to the device, therefore no interaction between the hypervisor and the VM is required. Only the assignment of a VF to a specific VM guest is done by the hypervisor before the guest is started.

Although SR-IOV is a powerful tool, there is no accelerator available which supports it, thereby requiring workarounds like the delegation methods introduced before. GPU vendor NVIDIA offers devices supporting SR-IOV, referred to as the GRID^{TM} series [NV13], however, these are sold as GPUs and not accelerators. Information on these products is scarce, so it can not be judged whether they would qualify as accelerators in professional environments. Compared to accelerators of the Tesla series, these devices are also less attractive from an economic perspective.

3 Concepts for efficient accelerator partitioning

In this section a concept for hardware supported accelerator partitioning based on SR-IOV is introduced. Before, the scope and definitions of the SR-IOV specification are clarified in the following. The SR-IOV specification of the PCI-SIG only defines

- how the PCIe configuration space of an SR-IOV capable device has to be configured,
- which features a device has to support in order to be SR-IOV capable and
- how VFs are implemented and managed, imposing some limitations device designs.

How device virtualization is actually realized depends on the device's design and functioning. It is therefore up to the designer to analyze the hardware to be virtualized and determine components which need to be available per VF or redesigned. Such an analysis will be conducted in this chapter.

Before performing the analysis, the basic considerations of the concept to be introduced are presented. The fundamental idea of the concept is to map every accelerator partition to one VF. Therefore a dedicated virtual device is available for each partition which can be assigned to a VM guest or regular application. By mapping partitions to VFs they are furthermore known to the accelerator's different subsystems which can use the added information to enforce SLAs like bandwidth guarantees. Partition setups are not static but can be changed when the corresponding VFs are not assigned to a VM, so the resources mapped by each partition can change over time.

In the remainder of this section state of the art accelerator architectures are analyzed. The Intel Xeon Phi and NVIDIA CUDA capable GPUs are used as representatives as they are both well documented and widely used. The analysis is done to determine components which have to be duplicated or redesigned to create an accelerator using SR-IOV to implement hardware partitioning support. Further details of different accelerator architectures are introduced as necessary to illustrate the proposed decisions and clarify possible limitations. For the sake of clarity, this section is divided into several subsections, each of which covers a distinct part of the concept.

3.1 Partitioning computational resources

When partitioning an accelerator, the first thing that comes to mind is partitioning its computational resources which are implemented using distinct functional units referred to as compute elements hereafter. Usually there is some granularity down to which partitioning makes sense as usually compute elements are grouped in some way, referred to as compute groups in the following. The reason for grouping compute elements is to improve the ratio of logic implementing control functionality to logic implementing compute elements. For NVIDIA accelerators, compute groups are called Streaming Multiprocessors (SMs) which consist of several compute elements called Streaming Processors (SPs) [Li08]. Besides the SPs, each SM also contains register files, instruction pointers as well as local memory for fast SP communication, data caching and data exchange [Li08]. The Intel Xeon Phi is structured differently as it consists of up to 61 P54C Pentium cores, which each comprise a 512 Bit SIMD vector unit, register file and cache [Ra13]. Compute groups and compute elements are harder to define here, as the processor cores are multi threading capable. Besides compute group limitations, a further aspect is important when deciding how to partition the compute elements of an accelerator. It has been noted before that cache interference effects can severely impact performance. As caches - at least L1 caches - are typically part of the compute groups, each partition should always be mapped to the same compute groups to avoid interference with other partitions. Each VF hence requires information about its assigned compute groups.

The operation of compute groups is controlled by an application running on the host machine or in the VM guest. Commands are sent to the accelerator by means of one or more command queues [CK09, Ne13] from which they are dispatched by the accelerator control logic and executed. Besides the command queue, mechanisms exist to communicate results and completions back to the controlling application [Ne13]. Several of these communication and command facilities may already exist for every recent accelerator in case it can be used concurrently by several applications or threads [Ne13]. For the proposed concept the communication and command facilities have to be linked to a VF by some means. Furthermore a mechanism is required which allows commands to be independently dispatched to the different compute group partitions. This suggests an individual command dispatch unit per partition and hence VF.

3.2 Partitioning memory resources

Accelerators usually have very wide memory interfaces and are equipped with fast memory chips to limit the memory bottle neck. The memory interface is typically realized using several memory controllers (c.f. [Li08, Ra13]). Therefore a trivial approach for memory partitioning could be to assign each VF one or more of the memory controllers. Although viable, this is not the best solution as it ties the memory bandwidth allocated to a VF to the amount of memory allocated to it. For example, if a VF with three quarters of the accelerator's memory bandwidth is requested, it would have to be assigned three quarters of the device's memory which it may not require. Furthermore bandwidth would be wasted in times when the VF is not running any calculations or a running program is not accessing memory. The approach envisioned for this concept is to equip memory controllers with a scheduler and provide a distinct request queue for every VF. As long as all request queues are filled, the scheduler arbitrates the queues such that memory bandwidth is shared between the VFs as requested. In case some of the queues are empty the bandwidth is shared only between those queues which contain requests so they can exceed their assigned bandwidth. Each queue set is bound to one scheduler. One part of the evaluation is to determine whether several schedulers are required or beneficial to improve performance.

3.3 Partitioning system interface resources

All currently available accelerators are co-processors which are connected to the host system via PCIe. This interface allows the software running on the host system (which the software running in a VM guest is considered a part of) to access accelerator memory. It is also possible, but not necessary, to provide access to host memory to the accelerator. Typically Direct Memory Access (DMA) automatons are integrated into accelerators to leverage the bandwidth of data exchange [Ra13]. DMA automatons are commonly controlled using command queues into which software inserts requests which the automaton serves in the order of their arrival. Different strategies are possible, e.g. to optimize the latency of short transfers, which might involve several distinct queues. However, this can be neglected for the general discussion. As data exchange between host and accelerator can

have a significant impact on program run time, a mechanism to partition system interface bandwidth is important.

To be able to make SLAs on system interface bandwidth for an accelerator partition, it is necessary to duplicate DMA command queues for each VF. The DMA automaton has to be redesigned to incorporate a scheduler which arbitrates the different DMA command queues. Several operations can be interleaved this way to share the available bandwidth. This assures that each VF has its assigned bandwidth available while being allowed to exceed it if other VFs do not fully utilize their assigned bandwidths.

3.4 Security concerns

It has been mentioned before that there are security concerns when using accelerators in multi tenant environments [Ma14]. One reason for such concerns is that accelerator memory is not overwritten when it is allocated, i.e. it still holds data of the process it was previously allocated to [BKS13]. While this is of minor concern in graphics applications and when using private accelerators, it becomes a problem when sensitive data is processed in a multi tenant environment. Sensitive data can leak to other users this way and maleficent users could even actively try to gather foreign data. As of now, the only option when processing sensitive data is to explicitly overwrite allocated memory before it is released. This could easily be automated by making the run time system of the accelerator overwrite memory when it is released. Although this sounds like a good fix at first, it creates unnecessary overhead, especially when data is allocated and released often. Another potential vulnerability is what can be referred to as volatile memory. Volatile memory in this context refers to memory which is not actively allocated by the user but implicitly by the run time environment, e.g. for passing function parameters. The most stringent solution appears to be strictly separating the memory regions assigned to the individual partitions. Separation should be enforced by a hardware protection mechanism, which is also desirable due to further factors which are discussed in the following paragraph.

Besides the security problem of relict data in multi tenant environments, another critical point is the protection of memory partitions against accesses from applications running in other partitions. Memory protection is possible on the Xeon Phi as it integrates a Memory Management Unit (MMU) [Ra13]. While [BKS13] states that accessing other processes' data is not possible using the CUDA API, this does not preclude that it could be accessed using lower level techniques. To provide a sound protection of sensitive data, it is part of this concept to research as to which extent memory management or protection is required and meaningful in the context of accelerator partitioning.

3.5 Synopsis: Proposed accelerator architecture

Summarizing all aforementioned aspects an architectural view of an accelerator with hardware partitioning support can be designed. Figure 3 shows an overview of the conceptual accelerator design containing all components discussed in the previous sections. Some

components, like the memory and system interface controllers, require redesign while there are others, like the memory protection mechanism, which are newly introduced. Further components, like command queues, have to be duplicated as they are required to be unique per VF and hence partition.

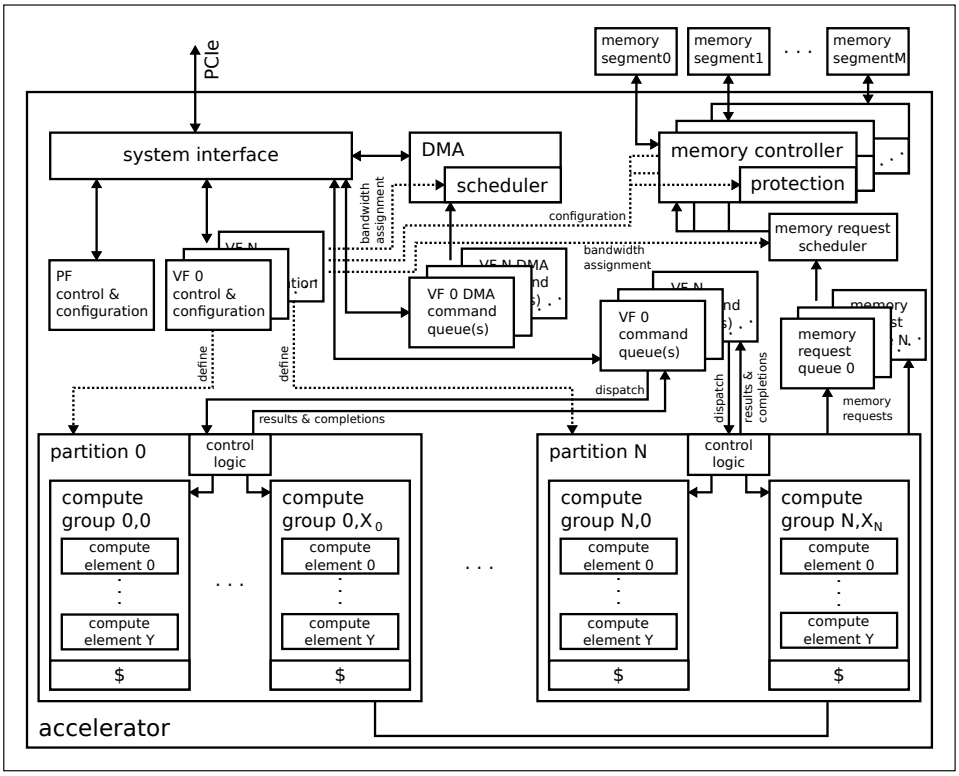


Figure 3: Overview of the proposed accelerator design concept.

4 Research methodology

Research will be conducted using a simplified model of a GPU based accelerator. A PCIe connected Field Programmable Gate Array (FPGA) will serve as an experimentation platform, SystemC can be used for early simulations and performance estimations. Developed variants of the concept can be implemented and evaluated using the experimentation platform, thereby giving a first impression of each variant's hardware costs and possible overhead. If time permits, the final experimental design could be synthesized using a state of the art standard cell library to evaluate the hardware costs of a real implementation. As of writing, work on the experimentation platform is still at an early stage. Currently an SR-IOV capable PCIe core is being developed as only the most recent FPGAs – which are only available on high priced boards – have integrated SR-IOV support. An existing open source PCIe core supporting SR-IOV could not be found.

The concept introduced in this paper has been developed by analyzing the general requirements of hardware partitioning support for accelerators. Several details are not fixed at the time of writing and often a multitude of options exists. Examples of such details of the design are the realization of the memory protection mechanism discussed in Section 3.4 or the different scheduling mechanisms. When refining the concept, design space evaluations are employed to explore and compare the available options and focus on the most suitable and promising ones.

One complicated point is the evaluation of the final design as a direct comparison with commercially available accelerators is illusory. The performance of commercial accelerators is impossible to reach within this project, the evaluation therefore has to be performed using indirect metrics. Currently the following evaluation concepts are considered:

- Comparisons of the performance achieved when the accelerator is partitioned to the baseline performance of the unpartitioned device. Performance ratios for different partitioning scenarios provide an indirect metric which can be used to compare the developed architecture to existing ones.
- Comparison to other solutions like those introduced in Section 2. Aspects like latency and bandwidth implications are of core interest. This comparison is considered important but depends on the availability and adaptability of the respective middleware.
- Interpolation to project the results to a possible silicon implementation. Results from interpolation can be used for comparing performance metrics to other accelerators.

5 Integration into the context of the dissertation

Commercially available accelerators, especially GPU based ones, have not evolved far from their roots. The traditional concept of one GPU for one computer still strongly affects accelerator usage patterns. This legacy concept impacts the performance of these powerful computing devices when integrating them into both existing and emerging usage scenarios. In the dissertation, two scenarios are focused:

- cluster computing as the predominant usage pattern in High Performance Computing (HPC) and
- device virtualization as an emerging usage pattern in cloud environments and an option to cut costs for machines with irregular utilization.

The goal is to propose an accelerator hardware architecture which meets the demands of the two aforementioned scenarios. It is expected that additional hardware support can greatly improve performance in both scenarios. However, it is obvious that they have quite different demands. Cluster computing is expected to benefit strongly from an optimized communication between several accelerators which are likely to be installed in

different machines. While efforts by NVIDIA were already made (marketed as GPUDirect RDMA [Sh11]), they are only applicable to one kind of interconnect. My goal in this context is to develop general concepts for transparently and efficiently integrating accelerator controlled data exchange via an arbitrary network. Device virtualization on the other hand requires the accelerator to be concurrently usable from several VMs. Mechanisms to realize this are available as discussed in Section 2, but a hardware solution is considered advantageous as stated in Section 3.

Important constraints for the concepts developed in the course of the dissertation are resource and overall efficiency. The additional features should have low hardware costs to limit their impact on chip costs. Furthermore they must not adversely affect regular operation of the accelerator, i.e. they should not have a negative impact on local and unpartitioned operation.

6 Acknowledgments

The work on this concept was in part supported by a grant by the federal state of Saxony.

References

[BKS13] Breß, S.; Kiltz, S.; Schäler, M.: Forensics on GPU Coprocessing in Databases – Research Challenges, First Experiments, and Countermeasures. In: Proceedings of the Workshop on Databases in Biometrics, Forensics and Security Applications (DBforBFS) at BTW 2013. Gesellschaft für Informatik, pp. 115–129, 2013.

[CK09] Croix, J. F.; Khatri, S. P.: Introduction to GPU Programming for EDA. In: Proceedings of the 2009 International Conference on Computer-Aided Design. ICCAD '09, ACM, New York, NY, USA, pp. 276–280, 2009.

[DS09] Dowty, M.; Sugerman, J.: GPU Virtualization on VMware's Hosted I/O Architecture. SIGOPS Oper. Syst. Rev., 43(3):73–82, July 2009.

[Du10] Duato, J. et al.: An Efficient Implementation of GPU Virtualization in High Performance Clusters. In (Lin, H. et al., eds): Euro-Par 2009 - Parallel Processing Workshops, volume 6043 of Lecture Notes in Computer Science, pp. 385–394. Springer Berlin Heidelberg, 2010.

[Go13] Gottschlag, M. et al.: LoGV: Low-Overhead GPGPU Virtualization. In: High Performance Computing and Communications 2013 IEEE International Conference on Embedded and Ubiquitous Computing (HPCC_EUC), 2013 IEEE 10th International Conference on. pp. 1721–1726, Nov 2013.

[GWO14] Grewe, D.; Wang, Z.; O'Boyle, M. F. P.: OpenCL Task Partitioning in the Presence of GPU Contention. In (Caşcaval, C.; Montesinos, P., eds): Languages and Compilers for Parallel Computing, volume 8664 of Lecture Notes in Computer Science, pp. 87–101. Springer International Publishing, 2014.

[Ha04] Harris, M.: Fast fluid dynamics simulation on the GPU. GPU gems, 1:637–665, 2004.

[Li08] Lindholm, E. et al.: NVIDIA Tesla: A Unified Graphics and Computing Architecture. Micro, IEEE, 28(2):39–55, March 2008.

[Lu08] Luebke, D.: CUDA: Scalable parallel programming for high-performance scientific computing. In: Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on. pp. 836–838, May 2008.

[Ma07] Manavski, S. A.: CUDA Compatible GPU as an Efficient Hardware Accelerator for AES Cryptography. In: Signal Processing and Communications, 2007. ICSPC 2007. IEEE International Conference on. pp. 65–68, Nov 2007.

[Ma14] Maurice, C. et al.: Confidentiality Issues on a GPU in a Virtualized Environment. In (Christin, N.; Safavi-Naini, R., eds): Financial Cryptography and Data Security, volume 8437 of Lecture Notes in Computer Science, pp. 119–135. Springer Berlin Heidelberg, 2014.

[Mo12] Montella, R. et al.: A General-Purpose Virtualization Service for HPC on Cloud Computing: An Application to GPUs. In (Wyrzykowski, R. et al., eds): Parallel Processing and Applied Mathematics, volume 7203 of Lecture Notes in Computer Science, pp. 740–749. Springer Berlin Heidelberg, 2012.

[Ne13] Newburn, C.J. et al.: Offload Compiler Runtime for the Intel® Xeon Phi™ Coprocessor. In: Parallel and Distributed Processing Symposium Workshops PhD Forum (IPDPSW), 2013 IEEE 27th International. pp. 1213–1225, May 2013.

[NV13] NVIDIA GRID™ GPU acceleration for virtualization. GPU Technology Conference, 2013.

[PC10] PCI Express® Base Specification Revision 3.0. Technical Specification, PCI-SIG, November 2010.

[Ra13] Rahman, R.: Intel® Xeon Phi™ Coprocessor Architecture and Tools. Apress, New York, 2013.

[Sh11] Shainer, G. et al.: The development of Mellanox/NVIDIA GPUDirect over InfiniBand – a new model for GPU to GPU communications. Computer Science - Research and Development, 26(3-4):267–273, 2011.

[Si10] Single Root I/O Virtualization and Sharing Specification Revision 1.1. Technical Specification, PCI-SIG, January 2010.

[TFJ94] Temam, O.; Fricker, C.; Jalby, W.: Cache Interference Phenomena. SIGMETRICS Performance Evaluation Review, 22(1):261–271, May 1994.

[Va08] Vaquero, L. M. et al.: A Break in the Clouds: Towards a Cloud Definition. SIGCOMM Comput. Commun. Rev., 39(1):50–55, December 2008.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

From Sensibility to Responsibility: The Impact of System Software on Ecological and Economical Sustainability of Computing Systems

Timo Hönig [1]

Abstract: From an ecological and economic perspective, saving energy is an important factor for the design of today's computing systems. However, building energy-efficient systems is a challenging task for system designers. It requires exact knowledge about the system's components (i.e., software and hardware components) and it is not immediately obvious *how* the energy demand of a system can be improved. This paper presents energy-aware programming techniques that help developers at optimizing the energy demand of software. Such techniques address *sensibility* aspects (i.e., energy awareness) and provide optimization suggestions to developers. The paper further discusses *responsibility* aspects that emerge from the advent of energy-aware programming techniques.

Keywords: Energy-Aware Programming, Energy-Aware Systems, System-Software Design, Sustainability of Computing Systems.

1 Introduction

Energy has become the most precious and critical resource of computing systems for all types of computing systems: from smallest-sized embedded systems up to large-scale data-center systems. For different kinds of computing systems, energy resources need to be handled with caution for different reasons. Embedded systems (e.g., smart dust [KKP99]) suffer from limited energy resources. It is necessary to address restricted energy resources, which arise from limited energy-storage capacities (e.g., supercapacitors, batteries), and the volatile availability of energy [Ka07]. Thus, embedded systems need to conserve available energy resources in the most efficient manner in order to avoid operational disruptions and to proactively prevent system failures caused by depletion of energy resources. Large-scale data-center systems (i.e., high-performance computing systems, cloud-computing systems) enjoy the practically unlimited availability of energy resources. However, large-scale systems commonly suffer from excessive consumption of energy (e.g., caused by operation at maximum performance) and entailing unwanted side effects (e.g., thermal overheating [Mo05], operating costs). Therefore, future large-scale and many-core systems [Es11] must be designed according to their energy characteristics and need to cautiously schedule activities to avoid system failures caused by excessive energy demand.

The effective energy-efficiency of any computing system depends on the characteristics of its hardware *and* software components. Hardware components with low-energy characteristics are optimized at CMOS level [Ja12] which reduce the energy demand of the

[1] Friedrich–Alexander–Universität Erlangen–Nürnberg (FAU), Department of Computer Science, Distributed Systems and Operating Systems, Martensstr. 1, D-91058 Erlangen

transistor circuit, for example, by implementing energy-saving modes and semiconductor scaling. However, hardware components only provide the basis for the energy-efficient operation of computing systems. It is necessary for the system software to operate in an energy-aware manner on the hardware, for example, by avoiding unnecessary system activities and exploiting hardware energy-saving features. Unfortunately, it is a difficult task for system-software designers to establish energy awareness of software components. This circumstance is caused by missing tool support, as programmers often lack energy-analysis tools that would help to identify energy hotspots of program code.

To reduce the energy demand of a computing system it is necessary to optimize the system and application software running on the system. First, the software itself needs to be optimized statically *before* run time (i.e., at the time of program-code creation) at different levels of abstraction (e.g., high-level language, assembly). Second, the software needs to be optimized dynamically *at* run time (i.e., at the time of program-code execution).

Profound tool support is an essential measure to help programmers at designing energy-aware computing systems, which exploit energy resources in the most efficient manner. For one, tool support actively assists programmers at designing energy-aware system software by integrating automated code analysis techniques. For another, tool support bridges the gap between developers and the operating system: important knowledge about the intent of the program code is transferred from the program design and programming phase (before run time) to the program execution phase (at run time) of the software.

The paper is divided into two parts. The first part of the paper (Sections 2 and 3) presents challenges in energy-aware system design and proposes concepts to address these challenges. This includes energy-aware programming techniques that establish the *sensibility* of energy demand. The second part of the paper (Section 4) analyzes the resulting impact on the *responsibility* of programmers with regards to the ecological and economical sustainability of system and application software. In more detail, the paper is structured as follows. Section 2 introduces energy-aware computing systems, presents related work, and discusses distinct energy-optimization methods. Energy-aware programming concepts and proactive system design techniques are presented in Section 3. The paper further discusses the consequences of establishing concepts for proactive energy-aware system design in Section 4. Section 5 outlines how the presented concepts are embedded into the author's doctoral thesis. Section 6 summarizes the proposed approaches and concludes the work.

2 Energy-Aware Systems

Energy-aware systems are computing systems which are conscious regarding the energy impact of individual operations. This energy awareness allows systems to economize available energy resources. The motivation for economizing energy varies and depends on the type of computing system: energy savings lead to extended battery life (i.e., embedded systems), reduced operating costs (i.e., desktop and server systems), and lower energy bills (i.e., large-scale systems).

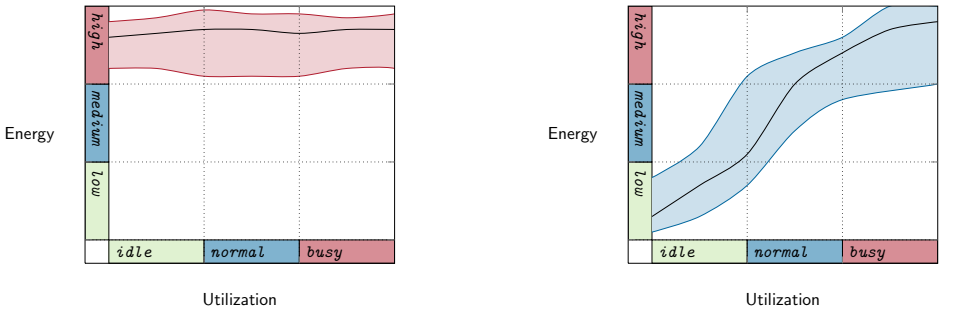


Fig. 1: Schematic representation of energy demand as a function of utilization: energy demand of an energy-unproportional system (left graph) vs. energy-proportional system (right graph).

2.1 Background

The implementation of an energy-aware system requires two main components. First, the system hardware needs to be energy proportional. Energy proportionality is achieved by implementing energy-saving features at hardware level which allow the system software (i.e., operating system) to influence the behavior of hardware components (e.g., activation and deactivation of functional features, switching of operating modes). Second, the system software needs to implement strategies to make efficient use of available energy-saving features at hardware level.

The hardware components implement low-power modes and sleep states at CMOS level and need to expose corresponding control interfaces that are used by the system software (i.e., operating system). Early computer systems did not implement any of such energy-saving features at hardware level, which ruled out software measures to adjust the energy demand depending on the systems' utilization. Today's systems, however, allow fine-grained control of the energy-saving features implemented by the hardware. Figure 1 shows the energy demand as a function of utilization for energy-unproportional systems (left graph) and energy-proportional systems (right graph).

From a system-software point of view, energy demand of computing systems can be decreased in two different ways. First, the program code of the system software can be optimized to exploit energy-saving features at hardware level in the most efficient manner. This includes device-specific low-power states which largely effect the static energy demand of the system. For example, this includes switching off underutilized hardware components such as wireless controllers. Second, software itself can be optimized to implement a required functionality in the most energy-efficient manner. This is achieved by structural optimizations of program code and the utilization of functional units at hardware level (e.g., floating-point units, graphics processing units).

2.2 Related Work

At system level, systems reduce their energy demand by exploiting device-specific energy-saving functions, which are implemented at hardware level by the individual system components (e.g., CPUs, wireless controllers [Ha10]). Today, energy-saving functions like

multi-level low-power modes and sleep states complement traditional approaches such as dynamic voltage and frequency scaling [CSP05]. On top of energy-saving primitives at hardware level, system software components commonly implement different strategies to increase the gain of energy savings. Traditionally, race-to-sleep strategies [DHKC09] have been proven to be effective measures to save energy. However, recent research [Hö14] shows that crawl-to-sleep strategies can be more energy-efficient.

At application level, software needs to be optimized statically (before run time) and dynamically (at run time). Recent proposals for energy-aware programming languages [Sa11] reduce the energy demand of computational resources required to execute the software by applying approximation techniques. Compiler optimizations [PHB15] at architecture level can not address energy-saving features at platform level (e.g., sleep states).

2.3 Static Energy Optimizations of Software

To implement *static optimizations* for reducing the energy demand of software before run time, developers need to optimize the program code at the time of program-code creation. Static optimization steps concern different levels of abstraction: First, program code needs to be altered in order to efficiently exploit available hardware functions (e.g., floating-point units, graphics processing units) and explicitly using available energy-saving functions (e.g., low-power modes, sleep modes) of the target platform. To apply static optimizations at level of the programming language, developers need to analyze the target hardware platform and adapt their program code according to the hardware characteristics. Second, different code transformations at lower language levels (e.g., intermediate language, machine language) than the originating high-level programming language apply code optimizations which influence the energy demand of the program code.

2.4 Dynamic Energy Optimizations of Software

To implement *dynamic optimizations* for reducing the energy demand of software at run time, the operating system must coordinate potentially disaccording interests (i.e., conflicting policies at task and system level) and the operating system continuously needs to execute the corresponding decision making process. However, the basis of this decision making process often lacks important information about the intended behavioral aspects of program code which currently executes on the computing system.

2.5 Problem Statement

The different energy-saving opportunities and the various static and dynamic approaches to optimize software for energy demand make it difficult for programmers to find the right set of measures to build energy-aware systems. Depending on hardware and software characteristics, programmers need to find the correct settings in a case-by-case decision process. This process is especially complicated as sufficient tool support often is missing. During the task of programming, developers commonly are left without orientation, as they have no feedback on the energy-demand of their programming decisions.

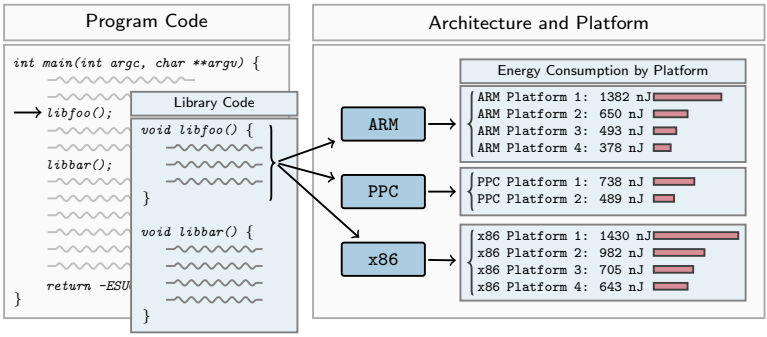


Fig. 2: Energy consumption of program code at function level: the individual energy consumption for executing a function depends on the hardware architecture and the hardware platform.

3 Energy-Aware Programming

Energy-aware programming techniques unify different concepts and increase the energy awareness of the developer during program-code creation. To initially create a basis of decision-making for the developer, energy-aware programming techniques provide feedback on the energy demand of program code. The techniques include code analyses that provide direct feedback on the energy demand of program code. As the energy demand differs for heterogeneous hardware architectures and platforms, analysis techniques must address this diversity, for example, by using virtualization technologies. The energy-function mapping in Figure 2 shows that energy demand for executing the program code of function `libfoo` differs on heterogeneous hardware architectures and platforms.

Energy demand of program code can be evaluated at different levels of abstraction. Programmers are most commonly interested in high-level energy analysis (i.e., energy analyses at function-level), as this allows them to reason about the effectiveness of code restructuring measures with regards to the energy efficiency. Depending on the analysis method, it requires often several intermediate steps to obtain the end result: energy-analysis methods evaluate the energy demand of the program code at other levels of abstraction (i.e., assembly level, basic-block level) and consolidate the results in a bottom-up approach.

As a consequence of energy-aware programming, programmers reveal *energy faults*[2] at design time and programmers can verify whether corresponding countermeasures (e.g., restructuring of program code) lead to the desired effect (i.e., reduction of energy demand). To implement a reliable chain of tool support for establishing energy-aware programming it is required to provide developers with energy consumption data of program code (Section 3.1), to assist at analyzing different code paths (Section 3.2), and to implement automation aspects to address the complexity of analyses (Section 3.3). The resulting infrastructure allows the implementation of analysis techniques, which automatically extract actual optimization suggestions for developers (Section 3.4).

[2] An *energy fault* is the root cause for unnecessarily high energy consumption that may result in a runtime *error* (deviation from target or actual) or even entails *failure* (breakdown).

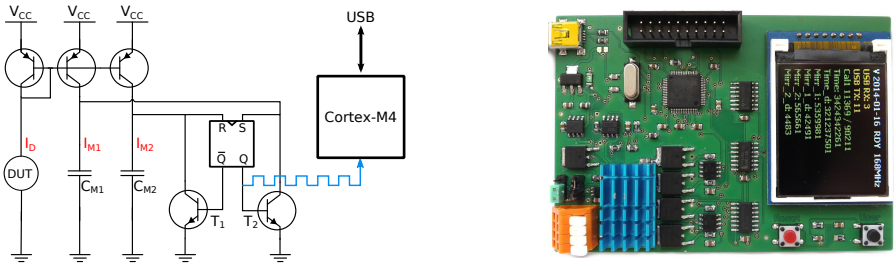


Fig. 3: The PEEK energy measurement device implements automated energy measurements and performs a current-to-frequency conversion for non-discrete measurements of energy demand.

3.1 Sensing Energy Demand

Energy-aware programming concepts rely on methods to sense and to quantify the energy demand of program code. This is either done by using energy-modeling [WB02, Hä12] or energy-measuring methods [TMW94, Fo08]. Model-based energy analyses are implemented by software components only. Therefore, programmers do not need to instrument their setup with additional hardware components (i.e., measurement equipment). As hardware complexity is steadily increasing, the creation of reliable and accurate energy models is a difficult task. For example, non-deterministic effects (i.e., cache misses, pipeline stalls) need to be considered by energy models. Results from model-based energy-analyses need to be verified periodically by control measurements to retain accuracy criteria. Measurement-based energy analyses use hardware instruments (e.g., multimeter, oscilloscope) to determine the energy demand of software during execution. In contrast to model-based energy analyses, measurement-based energy analyses are integrated into the setup of the system which is analyzed (device under test, DUT). As energy measurements follow a black-box approach, non-deterministic effects are included implicitly.

PEEK [Hö14] implements a non-discrete energy measurement device (i.e., an energy measurement device which does *not* sample) to avoid sampling limitations. The measurement device features microcontroller operation via USB, automatic calibration functions, and is capable of running automated energy measurements. The device implements a transistor circuit consisting of a current mirror and a flip-flop to implement a current-to-frequency conversion. Three PNP transistors mirror the input current I_{DUT} of the device under test to the mirrored currents I_{M1} and I_{M2}. The circuit operates as follows while the input current I_{DUT} is being drawn: controlled by an RS flip-flop, two capacitors (C_{M1} and C_{M2}) are being charged and discharged alternately. When the RS flip-flop outputs a logical 1 on the output Q, the path I_{M1} is pulled to ground via the transistor T_2. The path I_{M2}, however, is allowed to charge up the capacitor C_{M2}. Once C_{M2} reaches a voltage level, which the flip-flop recognizes as a logical 1, the flip-flop toggles its output Q. Now, the path I_{M2} is shorted to ground while path I_{M1} is charging up its capacitor C_{M1}. During each cycle, one capacitor charges while the other one is being discharged. The switching frequency of the output Q is directly proportional to the current I_{DUT}. The device uses the signal to determine the energy consumption of the DUT. Figure 3 shows an excerpt of the board schematics and the PEEK measurement device.

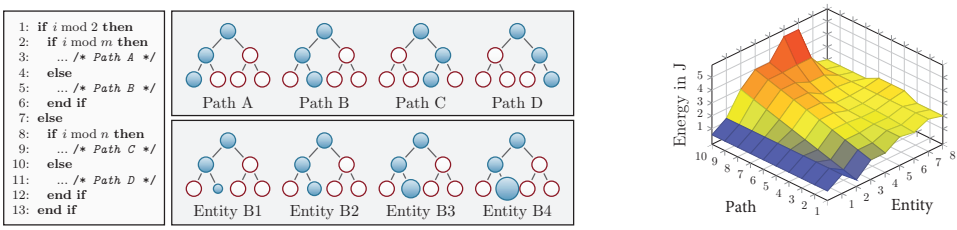


Fig. 4: The energy demand of program code at run time differs for various code paths and depends on the actual input data of the program (e.g., input parameters, processing data).

3.2 Combining Static and Dynamic Code Analysis

It is necessary to extract run-time characteristics to determine the energy demand of program code. Different program paths of an application may have entirely different energy demands. Furthermore, different program-path entities (i.e., identical program paths executed with different input data) again show distinct energy requirements. As an example, Figure 4 shows the excerpt of a program code that consists of four code paths. The value of the input parameter `i` determines which path is being executed at run time as visualized by the tree structure of the control flow graph. Moreover, the variable `i` influences the further proceeding of the code path, which results in path entities exposing distinct energy demands. The energy map in Figure 4 visualizes this.

To reason about the energy demand of individual program-code entities requires the use of static or dynamic code-analysis techniques. Static code-analysis techniques analyze the syntactical structure of code without actually executing it, whereas dynamic code analysis-techniques extract run-time information during the execution of the program code. Symbolic execution [Ki76, CDE08, Ca11] is an analysis technique that combines static and dynamic code-analysis techniques. Results from the code analysis are followed by a transformation step where energy demands are calculated on top of the analysis results.

The energy-aware programming technique SEEP [Hö12] uses symbolic execution techniques to automatically reveal possible program paths and to extract individual path constraints. Based on these path constraints, SEEP builds and executes binaries of the corresponding path entities. During this execution step SEEP extracts relevant run-time information (e.g., number and type of executed instructions) to calculate the energy demand of the individual path entities. Developers use such tool support to analyze whether specific programming decisions (e.g., restructuring of program code, using different program libraries) have a positive (or negative) impact on the energy demand.

3.3 Automated Analysis

Carrying out energy analysis for program code entails time-consuming operations. The amount of analysis work increases with the complexity of program code (i.e., growing number of code paths) and with the amount of potential target platforms that are considered by the developer. For this reason, energy-analysis tools need to be automated and run

unattended (i.e., without requiring manual interaction). Further, scalability aspects have to be considered and addressed at level of the tooling infrastructure of energy-aware programming techniques. For example, code-analysis methods use multiple threads or distributed computing systems to reduce the required analysis time.

3.4 Automated Energy Optimization

Energy-aware programming techniques which provide the infrastructure to sense energy demand and implement the execution of automated energy analyses are the basis for further extensions, for example, the automatic generation of energy-optimized program code.

Developers require hints and suggestions on *how* to actually improve their program code. A corresponding technique [Hö14] implements the generation of *energy optimization hints*. Energy optimization hints are being created during the code-analysis phase of energy-aware programming. If potential improvements for the program code of the developer are found, the infrastructure proposes a structural change to the original source code (e.g., as source-code patch). This step significantly improves the analysis process as the developer only needs to verify whether the suggested improvement introduces functional regressions.

FIGAROS [Hö15] improves energy-aware programming techniques by transferring knowledge from the developer to the run-time environment (i.e., operating system). The operating system kernel of FIGAROS is energy-aware and senses different system activities for energy analysis at run time. Future research work will integrate FIGAROS with the energy-aware programming kit PEEK [Hö14].

4 From Sensibility to Responsibility

The advent of energy-aware programming techniques improves the development of energy-efficient system software significantly. However, for developers the introduced *sensibility* of energy demand of program code implies a new *responsibility* as a direct consequence. Hence, this section discusses societal aspects tied to—and resulting from—the previously presented technical advances (i.e., energy awareness at programming level).

4.1 Ecological and Economic Impact

On the one hand, energy-aware programming techniques enable programmers to *evaluate*[3] quality aspects of their program code from an ecological point of view. For example, if a specific functionality can be implemented in different ways (i.e., different program structure) and with varying energy demand this leads to a new *value system* to evaluate program code. On the other hand, energy-aware programming techniques create an opportunity to address economical interests at the design time of computing systems. By saving prospective energy resources at design time, this builds new ways of optimizing systems from the bottom (i.e., system software) up to high-level program code (i.e., application software). The increased resource efficiency has an economic impact (i.e., monetary savings).

[3] The word *evaluate* is a compound word which originates from *ex-* (Latin for *out* or *from*) and *value* (Old French, Gallo-Romance dialect, for *value* → French: *évaluer* → English: *evaluate*).

4.2 Values and Sustainability

As software programmers reveal that the energy demand of program code is different among heterogeneous hardware platforms, this leads to decision situations that concern hardware aspects. For example, depending on the software of a computing system it can be ecologically reasonable to substitute the current hardware platform with a different hardware platform on which the program code has a smaller energy demand. However, such decisions need to consider further consequences. For example, sustainability aspects need to be respected. If a target hardware platform has already been obtained it is not always reasonable to move to a new hardware platform if the entailed demand of resources (e.g., production, energy, money) is higher than the savings of the platform substitution.

4.3 Verification and Certification

For large-scale projects among different partners (i.e., research groups, industry partners) it is necessary to consolidate energy demands of all subsystems. Energy-aware programming concepts are a complementary extension to existing project development processes (i.e., at design time) where individual components are verified by corresponding unit tests. For industrial use it is important to consider certification requirements (e.g., ISO 50001 [Mc10]).

5 Embedding into the Doctoral Thesis

The author's doctoral thesis proposes several distinct energy-aware programming concepts which are presented in extracts in Section 3. This work is basis for further research on reducing the energy demand of computing systems, including large-scale systems in distributed environments. The entailed responsibility (Section 4) is a logic consequence and needs to be considered by future work targeting at energy-aware system design. Research work contributing to this paper builds the science base of the author's doctoral thesis. The individual publications [Hö12, Hö14, Hö15] are presented with their abstract in the following three subsections.

5.1 SEEP: Exploiting Symbolic Execution for Energy-Aware Programming

The research [Hö12] presents the initial work on energy-aware programming and exploits the use of symbolic execution techniques to analyze program code for energy demand.

Abstract—In recent years, there has been a rapid evolution of energy-aware computing systems (e.g., mobile devices, wireless sensor nodes), as still rising system complexity and increasing user demands make energy a permanently scarce resource. While static and dynamic optimizations for energy-aware execution have been massively explored, writing energy-efficient programs in the first place has only received limited attention.

This paper proposes SEEP, a framework that exploits symbolic execution and platform-specific energy profiles to provide the basis for *energy-aware programming*. More specifically, the framework provides developers with information about the energy demand of

their code at hand, even for the invocation of library functions and in settings with multiple possibly strongly heterogeneous target platforms. This equips developers with the knowledge to take energy demand into account during the task of writing programs.

5.2 Proactive Energy-Aware Programming with PEEK

PEEK [Hö14] proposes a programming kit around the task of energy-aware programming. The work focuses on automation aspects and energy measurements.

Abstract—Optimization of application and system software for energy efficiency is of ecological, economical, and technical importance—and still challenging. Deficiency in adequate tooling support is a major issue. The few tools available (i.e., measurement instruments, energy profilers) have poorly conceived interfaces and their integration into widely used development processes is missing. This implies time-consuming, tedious measurements and profiling runs and aggravates, if not shoots down, the development of energy-efficient software.

We present PEEK, a systems approach to proactive *energy-aware programming*. PEEK fully automates energy measurement tasks and suggests program-code improvements at development time by providing automatically generated energy optimization hints. Our approach is based on a combined software and hardware infrastructure to automatically determine energy demand of program code and pinpoint energy faults, thereby integrating seamlessly into existing software development environments. As part of PEEK we have designed a lightweight, yet powerful electronic measuring device capable of taking automated, analog energy measurements. Results show an up to 8.4-fold speed-up of energy analysis when using PEEK, while the energy consumption of the analyzed code was improved by 25.3 %.

5.3 The FigarOS OS Kernel for Fine-Grained System-Level Energy Analysis

The research on FIGAROS [Hö15] takes energy-aware programming aspects to the operating-system level and it investigates how system kernels can be made self-conscious with regard to energy and how static and dynamic optimizations are addressed at operating-system level.

Abstract—Energy has become the most important operating resource for computing systems of all sizes—from embedded systems to large-scale high-performance computing systems. However, at system level, engineers remain challenged at efficiently handling energy as first-class operating system resource. The reasons for this are twofold: First, increasingly complex hardware circuits are inherently difficult to model which makes the creation of accurate energy models practically impossible. Second, available energy measurements at system level are coarse-grained and they are insufficient for fine-grained system level energy measurements of the operating system.

The current advent of power constrained many-core systems and the road ahead towards the era of dark silicon requires efficient energy control mechanisms in the system software layer. In this paper, we present FIGAROS, an operating system kernel which implements primitives required for fine-grained system-level energy analysis. Our implementation of FIGAROS orchestrates energy measurements at hardware level by a low-level system software infrastructure at kernel level.

6 Conclusions

This paper presents approaches for energy-aware programming to address ecological and economical sustainability aspects for computing systems. Energy-aware programming is an essential measure to establish sensibility for energy demand of program code right at the time of development. Assisted by corresponding tool support programmers are able to reason about their programming decisions with regard to energy demand. The introduction of sensibility for energy demand at programming level has several consequences for programmers—and their design decisions. Besides optimizing their program code by reducing its energy demand, programmers are now in position to control further ecological and economical aspects affected by system and application software.

References

[Ca11] Cadar, Cristian; Godefroid, Patrice; Khurshid, Sarfraz; Păsăreanu, Corina S; Sen, Koushik; Tillmann, Nikolai; Visser, Willem: Symbolic execution for software testing in practice: preliminary assessment. In: Proceedings of the 33rd International Conference on Software Engineering. IEEE, pp. 1066–1071, 2011.

[CDE08] Cadar, Cristian; Dunbar, Daniel; Engler, Dawson R: KLEE: unassisted and automatic generation of high-coverage tests for complex systems programs. In: Proceedings of the 12th Symposium on Operating Systems Design and Implementation. USENIX, pp. 209–224, 2008.

[CSP05] Choi, Kihwan; Soma, Ramakrishna; Pedram, Massoud: Fine-grained dynamic voltage and frequency scaling for precise energy and performance tradeoff based on the ratio of off-chip access to on-chip computation times. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 24(1):18–28, 2005.

[DHKC09] Dawson-Haggerty, Stephen; Krioukov, Andrew; Culler, David E: Power optimization: A reality check. Technical report, EECS Department, University of California, Berkeley, October 2009.

[Es11] Esmaeilzadeh, Hadi; Blem, Emily; St Amant, Renee; Sankaralingam, Karthikeyan; Burger, Doug: Dark silicon and the end of multicore scaling. In: Proceedings of the 38th International Symposium on Computer Architecture. ACM, pp. 365–376, 2011.

[Fo08] Fonseca, Rodrigo; Dutta, Prabal; Levis, Philip; Stoica, Ion: Quanto: Tracking energy in networked embedded systems. In: Proceedings of the 8th Symposium on Operating Systems Design and Implementation. USENIX, pp. 323–338, 2008.

[Ha10] Halperin, Daniel; Greenstein, Ben; Sheth, Anmol; Wetherall, David: Demystifying 802.11 n power consumption. In: Proceedings of the 2010 Workshop on Power-Aware Computing and Systems. USENIX, pp. 1–5, 2010.

[Hä12] Hähnel, Marcus; Döbel, Björn; Völp, Marcus; Härtig, Hermann: Measuring energy consumption for short code paths using RAPL. ACM SIGMETRICS Performance Evaluation Review, 40(3):13–17, 2012.

[Hö12] Hönig, Timo; Eibel, Christopher; Kapitza, Rüdiger; Schröder-Preikschat, Wolfgang: SEEP: Exploiting symbolic execution for energy-aware programming. ACM SIGOPS Operating System Review, 45(3):58–62, January 2012.

[Hö14] Hönig, Timo; Janker, Heiko; Eibel, Christopher; Mihelic, Oliver; Kapitza, Rüdiger; Schröder-Preikschat, Wolfgang: Proactive Energy-Aware Programming with PEEK. In: Proceedings of the 2014 Conference on Timely Results in Operating Systems. USENIX, pp. 1–14, 2014.

[Hö15] Hönig, Timo; Herzog, Benedict; Janker, Heiko; Schröder-Preikschat, Wolfgang: The FigarOS Operating System Kernel for Fine-Grained System-Level Energy Analysis. In: DAC Workshop on System-to-Silicon Performance Modeling and Analysis. ECSI, 2015.

[Ja12] Jan, C-H; Bhattacharya, U; Brain, R; Choi, S-J; Curello, G; Gupta, G; Hafez, W; Jang, M; Kang, M; Komeyli, K et al.: A 22nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high density SoC applications. In: Proceedings of the 2012 International Electron Devices Meeting. IEEE, pp. 1–4, 2012.

[Ka07] Kansal, Aman; Hsu, Jason; Zahedi, Sadaf; Srivastava, Mani B: Power management in energy harvesting sensor networks. ACM Transactions on Embedded Computing Systems, 6(4):32, 2007.

[Ki76] King, James C: Symbolic execution and program testing. Communications of the ACM, 19(7):385–394, 1976.

[KKP99] Kahn, Joseph M; Katz, Randy H; Pister, Kristofer SJ: Next century challenges: Mobile networking for "Smart Dust". In: Proceedings of the 5th International Conference on Mobile computing and Networking. ACM, pp. 271–278, 1999.

[Mc10] McKane, Aimee: Thinking globally: How ISO 50001 energy management can make industrial energy efficiency standard practice. Lawrence Berkeley National Laboratory, 2010.

[Mo05] Moore, Justin D; Chase, Jeffrey S; Ranganathan, Parthasarathy; Sharma, Ratnesh K: Making scheduling "cool": Temperature-aware workload placement in data centers. In: Proceedings of the 2005 Annual Technical Conference. USENIX, pp. 61–75, 2005.

[PHB15] Pallister, James; Hollis, Simon J; Bennett, Jeremy: Identifying compiler options to minimize energy consumption for embedded platforms. BCS The Computer Journal, 58(1):95–109, 2015.

[Sa11] Sampson, Adrian; Dietl, Werner; Fortuna, Emily; Gnanapragasam, Danushen; Ceze, Luis; Grossman, Dan: EnerJ: Approximate data types for safe and general low-power computation. In: Proceedings of the 32nd Conference on Programming Language Design and Implementation. ACM, pp. 164–174, 2011.

[TMW94] Tiwari, Vivek; Malik, Sharad; Wolfe, Andrew: Power analysis of embedded software: A first step towards software power minimization. IEEE Transactions on Very Large Scale Integration, 2(4):437–445, 1994.

[WB02] Weissel, Andreas; Bellosa, Frank: Process cruise control: Event-driven clock scaling for dynamic power management. In: Proceedings of the 2002 Conference on Compilers, Architecture, and Synthesis for Embedded Systems. ACM, pp. 238–246, 2002.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Towards Improving the Usability of Mobile ERP

A Model for Devising Adaptive Mobile UIs to Improve the Usability of Mobile ERP

Khalil Omar[1]

Abstract: Mobile ERP is an extension of ERP systems and many of business owners and CIOs remain reluctant to adopt this extension. The apparent reluctance stems from many challenges, most notably the usability when compared to conventional ERP packages. Usability challenges of mobile ERP could be classified based on every entity involved in these systems, namely mobile devices, mobile context, mobile connectivity, the end-user, and the back-end ERP system. The user interfaces of Mobile ERP directly affect its usability. Moreover, mobile ERP is used in variable contexts. Therefore, Adaptive user interfaces are proposed as a solution to address usability challenges and variable contexts issues by content, presentation and navigation adaptation. This research aims at providing a model for designing Adaptive Mobile User Interfaces to improve the usability of mobile ERP in variable contexts. Hence, this model will be used by mobile ERP developers to improve their product usability.

Keywords: Adaptive user interfaces; enterprise resource planning; mobile ERP; usability.

1 Introduction

Today, businesses have become increasingly reliant on Information Systems (IS) to support their business processes [MT12]. Enterprise applications are examples of using IS in businesses in order to become more flexible and productive [LL14]. One of the common examples of enterprise applications in practice are Enterprise Resource Planning (ERP) systems [LL14]. Many benefits are realized from adopting ERP systems in business, such as allowing real-time data streams between functional applications, integrating and processing large amounts of data [MT12].

The rapid technological advances in mobile computing improve modern businesses [SB12]. Enterprises started to migrate to the mobile strategy in order to meet changing requirements and expectations of their customers and business partners [DJ03]. Mobile enterprise term is used to describe a corporation or large organization that supports critical business functions and use of business applications via wireless mobile devices [GK14].

In practice, a significant number of enterprises employ a strategy of "best of breed" for their ERP systems to strive for a competitive advantage. Therefore, ERP vendors are scrambling to improve and extend their products features [GK14]. The migration to

[1] Carl von Ossietzky University of Oldenburg, Department of Computing Science, Ammerländer Heerstr 26129 Oldenburg, khalil.omar@uni-oldenburg.de

mobilize on-premise ERP systems is an urgent demand to execute enterprises processes anywhere and anytime to improve their Return on Investment (ROI) [Ca14].

Mobile ERP term was suggested by Willis in 2002, as an extension and future of ERP systems, that helps in solving the data capture problem [WW02]. Mobile ERP permits the user to take the ERP system anywhere [WW02]. Figure 1 shows that mobile ERP is the current wave of ERP systems [CS13].

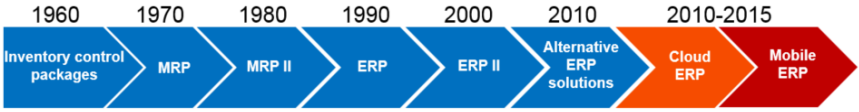


Fig. 1: Chronological evolution of ERP systems adapted from Cailean and Sharifi [CS13]

According to Dospinescu et al. *"mobile ERP is about having access to a software that allows a mobile device (portable computer, phone, Tablet PC, PDA) to be connected to the ERP system of an organization through a mobile net of communications and transmission of data GPRS/UMTS"* [Do08].

User interface layer is an important component of any software applications because it connects their end-users to the application functionalities. Therefore, software applications deployment could eventually fail due to a weak UI layer [Ak14]. ERP systems suffer from usability challenges due to their complex, rigid and bloated UIs [SW09],[Ak14]. Mobile ERP is an extension of ERP systems [WW02], thus this extension suffers from the same UIs complexity and rigidity, which directly affects the usability of mobile ERP.

Usability term was firstly coined in 1980s, to replace the term "user-friendly" [BKM92]. There are different models for usability that describe the usability attributes and factors. In 1994, Nielsen identified five attributes of usability, namely: efficiency, satisfaction, learnability, memorability, and errors [Ni94]. The International Organization for Standardization (ISO) defines usability as:

> *"The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specific context."* ISO 9241-11 (1998) [ISO98]

Usability is the most important quality attribute for mobile applications [NDA12]. According to Harrison et al., the previous models used for traditional desktop applications and the emergence of mobile devices added new challenges for usability. In fact, the previous models did not take into account these new challenges, such as cognitive load [HFD13].

In 2013, Harrison et al. proposed the usability model named People At the Centre of Mobile Application Development (PACMAD). This model considers the user, task and context of use are important factors to design usable applications, and identifies seven attributes that reflect the usability of an application:

1. **Effectiveness**: the user's ability to complete tasks in a specified context;
2. **Efficiency:** the user's ability to complete tasks with speed and accuracy;
3. **Satisfaction:** the user level of comfort and pleasantness through the use of software;
4. **Learnability:** the ease of learning to use the application;
5. **Memorability:** the ability of a user to remember how to use the application, if the application was not being used for a long period;
6. **Errors:** lower error rate, error recovering and avoiding catastrophic errors are proposed by Nielsen. PACMAD model extends the previous definition to include an evaluation of the errors while using mobile applications;
7. **Cognitive load:** the amount of cognitive processing required by the user to use the application, mobile application users are performing additional tasks, such as walking, unlike using traditional desktop applications.

Therefore, this research aims at improving the previous attributes of usability for mobile ERP.

2 Motivation

Many benefits are realized from adopting ERP systems in business, such as allowing real-time data streams between functional applications, integrating and processing large amounts of data [MT12]. ERP evaluation and selection processes need to consider the role of mobile applications in its overall business solution [Sc14]. According to Willis, mobile ERP is the future and another extension of ERP Systems that helps solving the data capture problem, it permits the user to take the ERP system with her/him [WW02].

Redshift Research Ltd. commissioned by Epicor Software Corporation recently conducted an online survey by interviewing 1500 business professionals in 10 countries. Many findings were revealed by this survey related to mobile ERP such as [Ep14]:

- *"65% perceive mobility as important to enable access to information and support communication for virtual workers.*
- *Only 1 in 2 have any form of remote access to their ERP systems.*
- *Only 25 percent can access their ERP via smartphones and tablet PC.*
- *In the future, 43% want ERP access via their smartphones, and 38 desire access via tablets."*

However, adopting ERP Systems over smartphones or tablet PCs requires a completely new designed user interfaces because of the limited space on the screen using touch-

based interactions [So14]. Furthermore, most ERP vendors provide generic mobile-to-enterprise data, which poorly display the standard ERP on mobile devices, which only reinforce complexity and frustration for user [Ri14]. Therefore, many CIOs are reluctant to invest more in mobility because of complexity concerns [Co13].

Usability challenges of mobile ERP could be classified based on every entity involved in this system, namely mobile devices, mobile context, mobile connectivity, the end-user, and the back-end ERP system [OM15].

- **Mobile Devices**[2]

  Smartphones and tablet PCs vary in their features and specifications. Actually, their capabilities are not comparable to desktop computers. The following is a brief explanation of mobile devices' features and its technologies that pose usability challenges for mobile ERP:

  - **Limited Screen Size**

    Mobile devices have a limited screen size compared to desktop computers. Therefore, there is a limitation on the information volume that can be displayed, and there is a huge possibility to lose its meaning.

    Van Baker, research vice president of Gartner Group, during a presentation to IT leaders in China in august 2014 said:

    "Enterprise application development teams use traditional practices to define and develop desktop applications; however, most don't work with mobile app development, due to device diversity, network connectivity and other mobile-specific considerations,"

    "The experience associated with mobile devices is significantly different from that of desktop devices, including shorter session lengths and limited presentation, due to screen size constraints that affect how mobile apps need to function" [Cy14]

    Adopting ERP systems on smartphones or tablet PCs require a completely new designed user interface because of the limited space on the screens and operation via touch screens [So14]. Accordingly, the limited screen sizes of mobile devices will adversely affect mobile ERP end-user satisfaction attribute.

  - **Limited Processing and Power Capabilities**

    Mobile devices are less powerful than desktop computers in respect to processing capability, memory and power [ZA05]. Therefore, some business functions and even ERP application features which were developed to facilitate the usage will be dispensed. As a result, this will directly impact

[2] This research is focusing on smartphone devices and tablet PCs

on the usability of mobile ERP.

- **Data Entry Methods**

  Todays, most of mobile devices use touch-based interactions (virtual keyboard). A recent study at Northern Illinois University found that the average typing speed on a virtual keyboard was found to be 25 words per minute, compared to 63 words on conventional and notebook keyboards [Ki14]. Another study revealed that virtual keyboards had higher discomfort ratings than conventional keyboards [KAJ12].

  We conclude that the following: the usability of mobile ERP will be hindered by using virtual keyboards due to its slow input speed (efficiency). Moreover, the probability of causing errors will be increased (effectiveness), and the end-user feeling will be discomforted (satisfaction).

- **Diversity of Mobile Operating Systems**

  End-users have strong personal connections to their mobile devices and mobile Operating Systems like Apple IOS, android, windows phone, etc. Therefore, this will affect their satisfaction if they were forced to use a particular mobile OS [An14].

- **Security**

  Security is a major concern for business owners when it comes to access ERP systems via mobile devices. In spite of device options that are available: bring-your-own-device (BYOD), choose-your-own-device (CYOD) and corporate-owned, personally enabled (COPE) devices, securing valuable data in ERP systems and devices from unauthorized access is compulsory. This is besides having a balancing strategy between security and usability without hindering the satisfaction of end-user.

- **Mobile Context**

Dey et al. defines context as:

> *"any information that can be used to characterize the situation of entities (i.e., whether a person, place, or object) that are considered relevant to the interaction between a user and an application, including the user and the application themselves"* [DAS01].

The interaction between the end-user of the mobile application, nearby people, objects, and environmental elements may distract end-user attention because the end-user will not be tied to a single location [ZA05]. Additionally, performing additional tasks while using mobile application (cognitive load) may cause distraction [HFD13]. Therefore, mobile ERP end-user may suffer from distraction that may lead to inaccuracy of data entry and failure in completing business process (effectiveness). Furthermore, it may slow down the business process

completeness (efficiency).

- **Mobile Connectivity**

  Mobile connectivity attributes such as bandwidth and reliability will affect the usability of mobile ERP in effectiveness, efficiency, and error in case of slow or loss of internet connection. On the contrary, the desktop computers have a stable connection to the back-end ERP systems.

- **The End-User**

  Todays, companies design and develop their products with user-oriented methods instead of technology-oriented methods [NDA12]. End-users with different levels of skill (e.g., novices, experts, etc.) and ages need to be considered when adopting ERP systems into mobile context.

  According to the European Commission report "the 2015 Ageing Report" [Eu14], people aged 65 and over will become a much larger share of population by 2060, the expected rise percentage is 18% to 28% of the population, and this implies to:

  > *"the EU would move from having four working-age people for every person aged over 65 years to only two working-age persons"*

  Elderly people usually have problems in the senses and movement and they will make up a large share of labour force in the near future [Eu14]. Therefore, mobile ERP vendors should consider this while dealing with usability.

- **Back-end ERP Systems**

  In recent years, several research works have revealed that ERP systems suffer from specific usability challenges [La14]. ERP systems integrate and process large amounts of data, therefore, their user interfaces are complex, rigid and suffer from poor usability [SW09]. Moreover, these systems have bloated user interfaces that affect end-user effectiveness [Ak14].

  Mobile ERP suffers from usability challenges which exists in ERP systems, because of its complex user interfaces. Furthermore, the topic of mobile ERP is still a young topic in practice and research [CS13]. Table 1 shows related works and user studies focusing on user interface aspects in ERP usability in stationary devices. We conclude that there is a lack of research works and empirical studies focusing on the usability of mobile ERP.

Author(s)	Year	Data sample	Methodology	System	Main aspects

Topi et al.[TLB05]	2005	9 users 1 non-user	in-depth interviews; semi-structured	unknown ERP (confidential)	• identification and access to information • transaction execution • system output • error support • terminology • system complexity
Singh & Wesson [SW09]	2009	3 experts	literature review; heuristic evaluation	SAP Business One	• navigation • presentation • task support • learnability • customization
Scholtz et al. [SCC10]	2010	21 users (students)	user study with questionnaires and diaries	SAP R/3	• navigation • presentation • task support • learnability • customization
Parks [Pa12]	2012	38 users	user study with talk-aloud and measurements	PeopleSoft	• UI complexity • task success • task time

Tab. 1: User studies focusing on user interface aspects in ERP usability [La14]

3 Problem Definition

Mobile ERP is the current wave of ERP systems [CS13], and this wave is currently a young topic in practice and research, therefore, more research works and studies need to be conducted.

User interface layer is an important component of any software applications because it connects their end-users to the application functionalities. Consequently, software applications deployment could fail due to a weak UI layer [Ak14]. ERP systems suffer from usability challenges due to their complex, rigid and bloated UIs [SW09], [Ak14].

Mobile ERP is an extension of ERP systems [WW02], Consequently, mobile ERP might suffer from the same ERP UIs complexity and rigidity. This in turn can directly affect the usability of mobile ERP.

Many research works have exploited the adaptive UIs to address some usability

challenges of applications software and their variable contexts. Adaptive UIs (AUIs) have the capability to be aware of its context-of-use and automatically react to the changes of this context in a continuous way. This is done by changing the UI presentation, contents, navigation or even behaviour [Fo10]. Therefore, Mobile ERP can benefit from adaptive UIs to improve its usability when it is utilized in variable contexts.

The development process of mobile applications can't show a high level of usability in adopting ERP system into mobile context because it is using the traditional methodologies (desktop applications) for developing mobile enterprises applications.

Therefore, this research aims at providing a model for designing Adaptive Mobile User Interfaces (AMUIs) and a prototype as a proof of concept to improve the usability of mobile ERP. This model will be used by mobile ERP developers to devise adaptive user interfaces that will improve the usability of mobile ERP.

4 Research Objectives and Questions

The main objective of this research is to improve the usability of mobile ERP by exploiting adaptive user interfaces. Hence, the main research question is:

> **How can adaptive user interfaces be exploited for improving the usability of mobile ERP?**

In order to achieve the main research objective, several sub-objectives need to be achieved in this research.

RO1) To identify the usability issues and challenges of mobile ERP. Accordingly, we conclude the following research question:

RQ1) What are the usability issues and challenges of mobile ERP?

RO2) To determine how AMUIs could be used to resolve the identified usability issues and challenges from RO1. Accordingly, we conclude the following research question:

RQ2) How AMUIs could be modelled and used to resolve the identified usability issues and challenges in RO1?

RO3) To design and develop a prototype as a proof of concept for mobile ERP developers. Accordingly, we conclude the following research question:

RQ3) How can we apply the identified AMUIs model in RO2 for mobile ERP?

RO4) to evaluate the developed prototype, in order to answer the following question:

RQ4) What are the improvements in the usability of mobile ERP after applying AMUIs model?

5 Research Methodology

This research employs design science research to achieve the main research objective by following the design science research guidelines, namely: design as an artifact, problem relevance, design evaluation, research contributions, research rigor, design as a search process, and communication of research [He04].

Starting with the business needs, to ensure that this research meets the goal of relevance, the researcher needs to identify the usability challenges and problems of mobile ERP in practice by conducting a usability evaluation questionnaire for mobile ERP end-users. In addition to that, a field study will be conducted to identify mobile ERP usability issues using the appropriate heuristics of usability evaluation for mobile applications.

To ensure research rigor, this research relies on existing knowledge bases and literature review within three main domains, namely: mobile usability, enterprise mobility, and human computer interaction. This will be done through an exhaustive review of the state of the art of adaptive UIs in literature and in practice besides evaluating and analysing the existing solutions.

Modelling the expected artifact this research which is the AMUIs model. This model will explain how to devise adaptive UIs for mobile ERP developers to improve the usability of their products. The AMUIs model contains the following interconnected sub-models based on AUIs types. These models need to be identified by the researcher:

- **Content adaptation model:** this model will explain how the UI of mobile ERP adapts its content (texts, labels, and images) by processing and delivering a personalized information for a specific end-user, task and dialog for a specific mobile context of use.
- **Presentation adaptation model:** this model will explain how the UI of mobile ERP presents the most appropriate presentation version (multimedia elements, interaction techniques, layout and graphical attributes) for a specific user, task and dialog for a specific mobile context of use.
- **Navigation adaptation model:** this model will explain how the UI of mobile ERP adapts its navigation links for a specific user to accomplish a specific task and dialog without feeling lost in a specific mobile context of use.

The previous models will be presented to the scientific community and conferences with the purpose of receiving feedback and constructive comments. Accordingly, these sub-models will be updated based on obtained feedback.

The AMUIs model will be implemented as a proof of concept, and the developed prototype will be evaluated using the field study evaluation method. This method is done by monitoring the use of the resulted artifact in multiple projects and observation of any improvements. The field study will use the appropriate evaluation heuristics, in order to identify any improvements in mobile ERP usability after applying AMUIs model.

6 Conclusion

Mobile ERP is the current wave of ERP systems, however, this wave suffers from usability challenges, which can be classified based on every entity involved in these systems. Furthermore, this wave is currently a young topic in practice and research.

Mobile ERP is an extension of ERP systems, therefore, this extension suffers from the same UIs complexity and rigidity, which negatively impact its usability. AUIs are proposed as a solution to improve the usability of mobile ERP. Therefore, this research aims at providing a model for devising AMUIs and a prototype as a proof of concept to improve the usability of mobile ERP. This model could be used by mobile ERP developers for devising adaptive user interfaces.

7 Acknowledgement

This research is supervised by Prof. Dr.-Ing. Jorge Marx Gómez, Carl von Ossietzky University of Oldenburg, Department of Computing Science Business Information Systems / Very Large Business Applications (VLBA).

References

[Ak14] Akiki, P.: Engineering Adaptive Model-Driven User Interfaces for Enterprise Applications, PhD thesis The Open University, 2014.

[An14] Ann All: 10 Keys to Creating a Mobile ERP Strategy. http://www.enterpriseappstoday.com/erp/10-keys-to-creating-a-mobile-erp-strategy.html, 11/2/2014.

[BKM92] Bevan, N.; Kirakowski, J.; Maissel, J.: What is Usability? In: Proceedings of the 4th International Conference on HCI, Stuttgart 1991. Elsevier, 1992.

[Ca14] Castellina, N.: Mobile ERP: Taking ERP ROI into Your Own Hands. Aberdeen Group, 2014.

[Co13] cbronline: 86% of businesses fail to utilise mobility - Computer Business Review. http://www.cbronline.com/news/mobile-and-tablets/86-of-businesses-are-failing-to-utilise-mobility-4142102, 11/6/2014.

[CS13] Căilean, D.; Sharifi, K.: Mobile ERP: A literature review on the concept of Mobile ERP systems, Master's Thesis in Informatics, Jönköping University, Sweden,

10/12/2013.

[Cy14] Cynthia Lee: Gartner Says Traditional Development Practices Will Fail for Mobile Apps. http://www.gartner.com/newsroom/id/2823619.

[DAS01] Dey, A. K.; Abowd, G. D.; Salber, D.: A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-aware Applications. In Hum.-Comput. Interact. 2, S. 97–166, 2001.

[DJ03] Dabkowski, A.; Jankowska, A. M.: Comprehensive framework for mobile ERP system. In: Proc. 14th Int. Workshop on Database and Expert Systems Applications,Prague 2003. IEEE Computer Society, Washington DC, S. 890–894, 2003.

[Do08] Dospinescu, O. et al.: Mobile enterprise resource planning: new technology horizons. In Communications of the IBIMA 11, S. 91–97, 2008.

[Ep14] Epicor Software Corporation: ERP and Mobile, Social, and Cloud – New Global Research Findings Revealed. http://www.epicor.com/Company/Pages/whats-next.aspx.

[Eu14] European Commission Economic and Financial Affairs European Economy Main series: The 2015 Ageing Report: Underlying Assumptions and Projection Methodologies, 2014.

[Fo10] Fonseca, José Manuel Cantera: Model-Based UI XG Final Report. http://www.w3.org/2005/Incubator/model-based-ui/XGR-mbui-20100504/, 5/17/2015.

[GK14] Gelogo, Y.; Kim, H.-K.: Mobile Integrated Enterprise Resource Planning System Architecture. In International Journal of Control and Automation 3/14, S. 379–388, 2014.

[He04] Alan R. Hevner et al.: Design Science in Information Systems Research. MIS Quarterly 1, S. 75–105, 2004.

[HFD13] Harrison, R.; Flood, D.; Duce, D.: Usability of mobile applications: Literature review and rationale for a new usability model. In Journal of Interaction Science 1, S. 1–16, 2013.

[ISO98] International Organization for Standardization: Ergonomic Requirements for Office Work with Visual Display Terminals. Part 11: Guidance on Usability. 9241-11, International Standards Organisation (ISO). 9241(11), 1998.

[Ki14] Kim, J. H. et al.: Differences in typing forces, muscle activity, comfort, and typing performance among virtual, notebook, and desktop keyboards. In Applied ergonomics 6, S. 1406–1413, 2014.

[KAJ12] Kim, J. H. et al.: Are there differences in muscle activity, subjective discomfort, and typing performance between virtual and conventional keyboards: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2012, S. 4764–4767.

[La14] Lambeck, C. et al.: (Re-)Evaluating User Interface Aspects in ERP Systems -- An Empirical User Study. In: System Sciences (HICSS), 2014 47th Hawaii

International Conference on. IEEE, S. 396-405, 2014.

[LL14] Laudon, K. C.; Laudon, J. P.: Management information systems. Managing the digital firm. Pearson, Global edition, 2014.

[MT12] Motiwalla, L.; Thompson, J.: Enterprise systems for management. Prentice Hall, Boston, 2012.

[NDA12] Nayebi, F.; Desharnais, J.-M.; Abran, A.: The state of the art of mobile application usability evaluation: CCECE, 2012, S. 1–4.

[Ni94] Nielsen, J.: Usability engineering. Morgan Kaufmann Publishers, San Francisco, Calif., 1994.

[OM15] Omar, K.; Marx Gómez, J.: Usability challenges for mobile ERP systems. http://www.erp-management.de/node/447.

[Pa12] Parks, N. E.: Testing & quantifying ERP usability: Proceedings of the 1st Annual conference on Research in information technology, 2012, S. 31–36.

[Ri14] rimini Street: Mobility Solutions for ERP. http://www.riministreet.com/Documents/Collateral/Rimini-Street-White-Paper-Mobility-Solutions-for-ERP.pdf, 11/1/2014.

[SB12] Stieglitz, S.; Brockmann, T.: Increasing organizational performance by transforming into a mobile enterprise. In MIS Quarterly Executive 11, S. 189–204, 2012.

[Sc14] Schneider, J.: Fast-moving trends in mobile ERP, CRM technology. http://www.fiercecio.com/story/fast-moving-trends-mobile-erp-crm-technology/2013-09-11, 11/1/2014.

[SCC10] Scholtz, B.; Cilliers, C.; Calitz, A.: Qualitative techniques for evaluating enterprise resource planning (ERP) user interfaces: Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, 2010, S. 284–293.

[So14] Sontow, K. et al.: ERP in Practice - User Satisfaction, Benefits & Prospects. http://www.trovarit.com/erp-praxis/studie-erp-praxis/erp-practice.html.

[SW09] Singh, A.; Wesson, J.: Evaluation criteria for assessing the usability of ERP systems. In: Proc. of the 2009 annual research conference of the South African Institute of Computer Scientists and Information Technologists. ACM, South Africa, S. 87-95, 2009.

[TLB05] Topi, H.; Lucas, W. T.; Babaian, T.: Identifying Usability Issues with an ERP Implementation: ICEIS, 2005, S. 128–133.

[WW02] Willis, T. H.; Willis-Brown, A. H.: Extending the value of ERP. In Industrial Management & Data Systems 1, S. 35–38, 2002.

[ZA05] Zhang, D.; Adipat, B.: Challenges, methodologies, and issues in the usability testing of mobile applications. In International Journal of Human-Computer Interaction 18.3, S. 293–308, 2005.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Vermeidung von IT Störungen unter Einsatz eines Kontrollsystems zur Durchführung von IT Changes

Arthur Vetter[1]

Abstract: In diesem Beitrag wird ein Forschungsvorhaben beschrieben, um Störungen in IT-Systemen und Service Level Agreement Verletzungen aufgrund von IT Changes zu verhindern. Dazu wird ein Kontrollsystem auf Basis von Change Plänen entwickelt, welches die aktive Kontrolle von Change Durchführungen und dessen Risiken ermöglichen und somit eine direkte Steuerung von Changes in Echtzeit unterstützen soll. Es soll eine Vorhersage der Change Durchführung erfolgen, indem verschiedene Parameter, wie beispielsweise die Fehlerwahrscheinlichkeit oder die voraussichtliche Dauer einer bestimmten Change Aktivität, während der Change Durchführung analysiert und für den weiteren Verlauf des Changes geschätzt werden. Zur Verbesserung der Entscheidungsunterstützung sollen Business Process Intelligence Methoden eingesetzt werden, um Wissen aus bereits durchgeführten Changes nutzen zu können.

Keywords: IT Service Management, Predictive Change Management, Business Process Intelligence, Business Process Management, Change Plans, Incidents, Risikomanagement

1 Einleitung

Der Einsatz von Informationstechnologien nimmt in Unternehmen seit Jahren stetig zu und hat für viele Unternehmen eine geschäftskritische Bedeutung erreicht [MHG15]. Es ist davon auszugehen, dass aktuelle Trends, wie Digitalisierung, Big Data, Cloud Computing und Industrie 4.0 die Bedeutung der IT für Unternehmen noch weiter erhöhen werden. Dies führt dazu, dass ein Ausfall von IT-Services, mit dem wichtige Geschäftsprozesse unterstützt werden, zu massiven negativen Auswirkungen für die Kunden führen kann, wie folgendes Beispiel verdeutlicht:

Am 21. April 2011 kam es in einem Rechenzentrum von Amazon zu einer Störung, die zu einem temporären Netzwerkausfall führte, wodurch die angebotenen Dienstleistungen von Amazon nicht mehr erreicht werden konnten. Die Störung hatte einen temporären und sogar dauerhaften Verlust von Kundendaten zur Folge. Grund der Störung war die fehlerhafte Durchführung eines geplanten Netzwerk-Upgrades. Für die Durchführung des Upgrades sollte der Datenverkehr über einen redundanten Router geleitet werden, jedoch war der Router nicht für die Datenlast ausgelegt, so dass es zu Netzwerkstörungen kam [HSK12, S.368].

Für Unternehmen bieten sich verschiedene Möglichkeiten ungeplante Downtimes ihrer Systeme und Services zu verhindern und somit die Ausfallsicherheit ihrer IT-Systeme zu

[1] Karlsruher Institut für Technologie, AIFB, Kaiserstraße 89, 76133 Karlsruhe, arthur.vetter@partner.kit.edu

gewährleisten. Die Ausfallsicherheit lässt sich beispielsweise durch den Aufbau redundanter IT-Systeme, das Abschließen von Service-Verträgen und ein Availability Management zur kontinuierlichen Überwachung der Verfügbarkeit der IT-Systeme erhöhen [Sc99, S.1]. Allerdings sind lediglich ca. 20% der Störungen in IT-Systemen auf technische Probleme zurückzuführen. Die restlichen 80% der Störungen werden durch menschliches Versagen verursacht [Sc99, S.1]. Diese 80% der Störungen lassen sich wiederum zur Hälfte unterteilen in Anwendungsfehler, die bei der Softwareentwicklung gemacht werden, und Durchführungsfehler, die bei der Einführung eines Changes in die bestehende Systemlandschaft, wie im Beispiel von Amazon, entstehen [Sc99, S.2].

Zur Vermeidung von Störungen, die auf Durchführungsfehler zurückzuführen sind, wird der Einsatz eines Change Managements empfohlen [Sc99, S.2], [BJK08, S.26]. Der De-Facto Standard im IT-Service Management, die IT Infrastructure Library (ITIL), beschreibt die einzelnen Aktivitäten, die im Rahmen eines Change Managements durchzuführen sind, um einen Change kontrolliert umzusetzen. Die genaue Ausgestaltung des Change Managements lässt ITIL jedoch offen. Keller et al. haben 2004 den Prototypen "CHAMPS" erstellt, mit dem auf Basis einer Configuration Management Database[2] (CMDB) automatisiert Change Pläne erstellt werden können [Ke04]. Seit dieser Arbeit sind weitere Arbeiten und Prototypen zur Erstellung von Change Plänen erstellt und veröffentlicht worden [Co09a], [TFL09]. Change Pläne können als Schritt-für-Schritt Anleitung für die Durchführung eines Changes interpretiert werden und bieten somit Potential zur Verringerung von Durchführungsfehlern. Allerdings berücksichtigen die bisher vorgestellten Systeme nicht die Kontrolle der eigentlichen Durchführung, d.h. wenn während der Durchführung eines Changes ein Fehler passiert und der Change daher nicht wie geplant zu Ende geführt werden kann bzw. es aufgrund des Fehlers zu Störungen im Betrieb kommt, wie am Beispiel von Amazon aufgezeigt. Im Dissertationsvorhaben soll eine Lösung entwickelt werden, die es ermöglicht, einen Change und dessen Risiken während der Umsetzung zu kontrollieren, um direkt proaktiv steuernd eingreifen zu können, so dass Störungen aufgrund von Durchführungsfehlern bereits während der Change Durchführung verhindert werden können. Die proaktive Steuerung soll dabei durch eine Vorhersagefunktion unterstützt werden.

2 Ziele der Arbeit und Untersuchungsverfahren

Das Sachziel dieser Arbeit ist die Entwicklung einer Methode, mit der die Kontrolle und Steuerung von Change Durchführungen und dessen Risiken möglich sind, um Störungen aufgrund von Durchführungsfehlern und dadurch auftretenden Service Verletzungen zu vermeiden. Dabei soll eine proaktive Steuerung während der Change Durchführung in Echtzeit realisiert werden. Zusätzlich zum vorgestellten Sachziel, sollen folgende

[2] Als CMDB wird eine Datenbank verstanden, in der Configuration Items (CI) und deren Beziehungen zueinander gepflegt werden. Configuration Items sind Komponenten, die zur Bereitstellung eines IT-Services benötigt werden, zum Beispiel ein Server, eine Datenbank etc. [BJK08, S.129].

Formalziele erreicht werden:

- Es soll dynamisch während der Change Durchführung die Erfolgswahrscheinlichkeit eines Changes und die voraussichtliche Dauer des Changes berechnet werden.
- Um zukünftige Prognosen zu verbessern, sollen Erfahrungen aus bereits durchgeführten Changes genutzt werden.

Zur Erreichung des Untersuchungsziels wird überprüft, in wie fern sich die Methodik der Geschäftsprozessmodellierung dazu eignet, die Durchführung von Change Plänen zu kontrollieren. Dazu sollen die Aktivitäten des Changes in Form eines Change Plans modelliert werden. Um die Anforderung der dynamischen Kontrolle und Steuerung der Change Durchführung zu erreichen, soll eine interaktive Simulation zum Einsatz kommen. Dabei soll im Change Plan der Zustand der einzelnen Aktivitäten angezeigt werden, z. B. wird aktuell durchgeführt, bereits erfolgreich/erfolgslos durchgeführt, noch nicht durchgeführt, etc. Auf dieser Basis kann zum Beispiel bei einer fehlerhaften Durchführung entschieden werden, ob die Aktivität nochmals durchgeführt werden soll, oder lieber ein Roll Back des gesamten Changes erfolgen soll. Um den aktuellen Zustand der Change Durchführung im Change Plan anzeigen zu können, ist eine Echtzeitüberwachung der durchzuführenden Aktivitäten und des Zustands des IT-Systems notwendig.

Das genaue Untersuchungsverfahren wird anhand folgender Fragestellungen beschrieben:

- *Welche Modellierungssprachen eignen sich zur Abbildung und Simulation von Change Plänen?*

  Durch Beantwortung dieser Frage sollen mögliche Modellierungssprachen identifiziert werden, die sich zum einen zur Abbildung von Change Plänen eignen und zum anderen für eine interaktive Simulation verwendet werden können, um den aktuellen Status der Change Durchführung anzuzeigen.

- *Welche Informationen werden benötigt, um die Durchführung eines Change Plans zu kontrollieren und aus welchen Quellen lassen sich diese extrahieren?*

  Zur Kontrolle der Change Durchführung werden Sensoren zur Ermittlung des aktuellen Durchführungsstatus benötigt. Für die Entwicklung der Sensoren sind die Informationen zu identifizieren, die benötigt werden um den aktuellen und zukünftigen Zustand der Change Durchführung und den Zustand des zu ändernden IT-Systems zu beschreiben. Nachdem die benötigten Informationen herausgearbeitet sind, können die Informationsquellen identifiziert werden, aus denen die Informationen extrahiert werden können, um sie auszuwerten. Bei dieser Fragestellung werden explizit auch Informationen aus der Vergangenheit untersucht, wie beispielsweise die benötigte Zeit zur Durchführung einzelner Aktivitäten, oder die Fehlerwahrscheinlichkeit einer Aktivität, um das Formalziel

zur Verbesserung der Prognosen zu erreichen.

- *Welche Möglichkeiten zur Messung der notwendigen Daten gibt es?*

  Zur Entwicklung der Sensoren ist zu untersuchen welche Methoden sich zur Extraktion der Daten eignen und inwieweit sich diese automatisiert messen lassen. Dabei soll insbesondere das Potenzial von Process Intelligence[3] Methoden untersucht werden, als auch Infrastruktur Monitoring Lösungen.

- *Wie lassen sich die gemessenen Daten auswerten und Vorhersagen erstellen, so dass eine proaktive Steuerung je nach Zustand der Change Durchführung möglich ist?*

  An dieser Stelle soll ein Konzept entwickelt werden, mit dem die vorgesehene proaktive Steuerung ermöglicht wird. Dabei wird untersucht, wie mithilfe der gewählten Modellierungssprache der aktuelle Zustand der Change Durchführung auf Basis der gemessenen Daten simuliert werden kann und weitere Parameter, wie zum Beispiel die voraussichtliche Restdauer des Changes, die Fehlerwahrscheinlichkeit, etc. berechnet und vorhergesagt werden können. Als Problemlösungsverfahren soll dabei die Quality-of-Service Aggregation zum Einsatz kommen. Zudem ist sowohl nach der gesamten Change Durchführung als auch nach festgelegten Durchführungsschritten während der Change Ausführung zu überprüfen, ob sich das System im geplanten Zustand befindet.

- *Wie ist die Architektur eines solchen Change Kontrollsystems zu gestalten?*

  Zur Beantwortung dieser Frage soll ein Konzept für eine Architektur zur Extraktion der benötigten Daten, der Kontrolle der Durchführung der Change Pläne und der automatischen Berechnung der verschiedenen Paramater zur Prognose der Change Durchführung entwickelt werden.

Zur Validierung des Ansatzes ist die Entwicklung eines Prototypen und dessen Anwendung in einer Fallstudie geplant. In der Fallstudie soll ein Upgrade eines bestehenden ERP Systems durchgeführt werden.

Für die Arbeit werden folgenden Annahmen getroffen:

- Im Unternehmen besteht eine CMDB, in der alle IT-Komponenten, die zur erfolgreichen Durchführung eines Changes benötigt werden, dokumentiert sind.
- Alle Änderungen an einem IT-System werden in Form von Logs dokumentiert und sind für Auswertungen extrahierbar.
- Es besteht vollständiges Wissen über die durchzuführenden Aktivitäten für einen Change.

[3] Process Intelligence umfasst die Anwendung verschiedener Business Intelligence Methoden zur Analyse, Vorhersage und Optimierung von Prozessen [Gr04, S.321 f.].

3 Aktueller Forschungsstand und Ausblick

In den letzten Jahren wurden verschiedene Arbeiten zum Thema IT Change Management mit unterschiedlichen Schwerpunkten veröffentlicht. In einigen Arbeiten werden Optimierungsmodelle hinsichtlich unterschiedlicher Aspekte wie dem Business Impact oder der Konfliktfreiheit von Changes untersucht, um negative Auswirkungen von Changes aufgrund von Fehlern bei der Einplanung eines Changes zu vermeiden [Re07], [Sa07], [SBL08], [Zi08], [HSK12], [Ha12], [Ha14].

Andere Arbeiten befassen sich mit der Erstellung von formalen, automatisierbaren Change Plänen, in denen die durchzuführenden Aktivitäten beschrieben sind [Ke04], [Co09a], [TFL09], [Co09b]. Ein Konzept zur Erweiterung dieser Change Pläne um Risiken, die ex post ausgewertet werden, wurde von Wickboldt et al. vorgestellt [Wi11]. Weitere Arbeiten beschäftigen sich mit Fehlersituationen bei Changes, um die Ursache eines Fehlers zu identifizieren [Sa13], bzw. einen entsprechenden Rollback durchzuführen, um das geänderte System wieder in einen konsistenten Zustand zu versetzen [Ma08].

Diese Arbeiten beschreiben verschiedene Aspekte des Change Managements, die vor der Durchführung eines Changes (automatische Erstellung von Change Plänen) oder ex post (Auswertung der Risiken und Fehler) zur Anwendung kommen. Mit diesem Promotionsvorhaben soll ein Beitrag zur Automatisierung des Change Management Prozesses geleistet werden, in dem die Change Durchführung selbst in Echtzeit kontrolliert und gesteuert wird. Zudem soll durch das Werkzeug die Überwachung verteilter Changes, z.B. über mehrere Systeme und Standorte vereinfacht werden.

Literaturverzeichnis

[BJK08] Bon, J. v.; Jong, A. d.; Kolthof, A.: Service Transition basierend auf ITIL V3: Ein Management Guide. In: Zaltbommel: Van Haren Publ (2008)

[Co09a] da Costa Cordeiro, W. L.; Machado, G. S.; Andreis, F. G.; dos Santos, A. D.; Both, C. B.; Gaspary, L. P.; Granville, L. Z.; Bartolini, C.; u. a.: ChangeLedge: Change design and planning in networked systems based on reuse of knowledge and automation. In: Computer Networks Bd. 53 (2009), Nr. 16, S. 2782–2799

[Co09b] da Costa Cordeiro, W. L.; Machado, G. S.; Andreis, F. G.; Wickboldt, J. A.; Lunardi, R. C.; dos Santos, A. D.; Both, C. B.; Gaspary, L. P.; u. a.: Changeminer: A solution for discovering it change templates from past execution traces. In: IEEE, 2009 — ISBN 1424434866, S. 97–104

[Gr04] Grigori, D.; Casati, F.; Castellanos, M.; Dayal, U.; Sayal, M.; Shan, M.-C.: Business process intelligence. In: Computers in Industry Bd. 53 (2004), S. 321–343

[HSK12] Hagen, S.; Seibold, M.; Kemper, A.: Efficient verification of IT change operations or: How we could have prevented Amazon's cloud outage. In: IEEE, 2012 — ISBN 1467302678, S. 368–376

[Ha12] Hagen, S.; da Costa Cordeiro, W. L.; Gaspary, L. P.; Granville, L. Z.; Seibold, M.; Kemper, A.: Planning in the large: efficient generation of IT change plans on large infrastructures. In: IEEE, 2012 — ISBN 1467331341, S. 108–116

[Ha14] Hagen, S.; da Costa Cordeiro, W. L.; Paschoal G., L.; Granville, L. Z.; Kemper, A.: Efficient Model Checking of IT Change Operations (2014)

[Ke04] Keller, A.; Hellerstein, J. L.; Wolf, J. L.; Wu, K.-L.; Krishnan, V.: The CHAMPS system: change management with planning and scheduling. In: IEEE, 2004 — ISBN 0780382307, S. 395–408

[MHG15] Maes, K.; Haes, S. D.; Grembergen, W. V.: Developing a Value Management Capability: A Literature Study and Exploratory Case Study. In: Information Systems Management Bd. 32 (2015), Nr. 2, S. 82–104

[Ma08] Machado, G. S.; Daitx, F. F.; da Costa Cordeiro, W. L.; Both, C. B.; Gaspary, L. P.; Granville, L. Z.; Bartolini, C.; Sahai, A.; u. a.: Enabling rollback support in IT change management systems. In: IEEE, 2008 — ISBN 1424420652, S. 347–354

[Re07] Rebouças, R.; Sauvé, J.; Moura, A.; Bartolini, C.; Trastour, D.: A decision support tool to optimize scheduling of IT changes. In: IEEE, 2007 — ISBN 1424407982, S. 343–352

[Sc99] Scott, Donna: Making smart investments to reduce unplanned downtime. In: Tactical Guidelines Research Note Note TG-07-4033, Gartner Group, Stamford, CT (1999)

[SBL08] Setzer, T.; Bhattacharya, K.; Ludwig, H.: Decision support for service transition management Enforce change scheduling by performing change risk and business impact analysis. In: IEEE, 2008 — ISBN 1424420652, S. 200–207

[Sa07] Sauvé, J. P.; Santos, R. A.; Almeida, R. R.; Moura, J. A. B.: On the risk exposure and priority determination of changes in IT service management. In: Managing Virtualization of Networks and Services: Springer, 2007 — ISBN 3540756930, S. 147–158

[Sa13] dos Santos, R. L.; Wickboldt, J. A.; Dalmazo, B. L.; Granville, L. Z.; Gaspary, L. P.; Lunardi, R. C.: Identifying the root cause of failures in IT changes: Novel strategies and trade-offs. In: IEEE, 2013 — ISBN 1467352292, S. 118–125

[TFL09] Trastour, D.; Fink, R.; Liu, F.: Changerefinery: Assisted refinement of high-level IT change requests. In: : IEEE, 2009 — ISBN 0769537421, S. 68–75

[Wi11] Wickboldt, J. A.; Bianchin, L. A.; Lunardi, R. C.; Granville, L. Z.; Gaspary, L. P.; Bartolini, C.: A framework for risk assessment based on analysis of historical information of workflow execution in IT systems. In: Computer Networks Bd. 55 (2011), Nr. 13, S. 2954–2975

[Zi08] Zia, L.; Diao, Y.; Rosu, D.; Ward, C.; Bhattacharya, K.: Optimizing change request scheduling in IT service management. In: IEEE, 2008 — ISBN 0769532837, S. 41–48

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Studierendenkonferenz Informatik 2015 (SKILL 2015)

Studierendenkonferenz Informatik 2015 (SKILL 2015)

Hervorgegangen aus der Studierendenkonferenz Informatik Leipzig (SKIL) findet in diesem Jahr bereits die zweite SKILL in Zusammenarbeit mit der Gesellschaft für Informatik e.V. im Rahmen der GI-Jahrestagung INFORMATIK 2015 statt. Sie hat sich damit als feste Größe zur Veröffentlichung studentischer Arbeiten im deutschsprachigen Raum etabliert. Dieses Jahr umfasst das Konferenzprogramm 13 Beiträge, die eine große Bandbreite verschiedener Themen der Informatik abdecken.

Die Mitglieder des Organisationskomitees der SKILL 2015 bedanken sich bei allen Gutachterinnen und Gutachtern sowie bei den Sponsoren Infineon Technologies AG[1] und serioussearch.de[2]. Unser Dank gilt nicht zuletzt allen Autorinnen und Autoren, die mit ihren qualitativ hochwertigen Beiträgen diese Konferenz erst ermöglicht haben. Auch in diesem Jahr versteht sich die Studierendenkonferenz Informatik als Plattform zum überregionalen fachlichen Austausch und zur intensiven Diskussion.

Wir freuen uns besonders, dass wir dieses Jahr eine große Anzahl an Professorinnen und Professoren aus verschiedenen Fachbereichen der Informatik zur Unterstützung der SKILL gewinnen konnten. Aus diesem Grund konnte jeder eingereichte Beitrag von mindestens drei ausgewiesenen Expertinnen und Experten begutachtet werden. Die Liste aller Gutachterinnen und Gutachter findet sich auf der Webseite der SKILL 2015[3].

Organisationskomitee der SKILL 2015

Johannes Schmidt, Universität Leipzig

Thomas Riechert, HTWK Leipzig

Agnes Koschmider, Karlsruher Institut für Technologie

Michael Becker, Universität Leipzig

Heinrich Theodor Vierhaus, BTU Cottbus-Senftenberg

Ludger Porada, Gesellschaft für Informatik e.V.

Robert Scheffler, Fachschaftsrat Informatik BTU Cottbus-Senftenberg

[1] http://www.infineon.com/careers
[2] https://www.serioussearch.de/about.html
[3] http://skill.informatik.uni-leipzig.de/blog/skill2015/programmkomitee

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Eine Smart Grid Architektur und Umsetzung zwecks Datenanalyse und Leistungsmanagement am Beispiel des Stromnetzes der OTH Regensburg

Susanne Kenner[1] und Raphael Thaler[2]

Abstract: Smart Grid, Smart Metering, Elektromobilität und Netzregulierung sind Schlagworte der Energiewende. Die Energieversorgung soll intelligenter werden. Vor diesem Hintergrund forscht die Ostbayerische Technische Hochschule (OTH) Regensburg in verschiedenen Projekten innerhalb eines campusweiten Smart Grids. In diesem Beitrag wird eine Software vorgestellt, die im Rahmen eines Forschungsprojekts zur Erfassung, Visualisierung und Analyse von Stromdaten des Smart Grids entwickelt wurde. Die Applikation und Systemarchitektur dienen u.a. als Grundlage für weitere Forschungen im Bereich des automatisierten Leistungsmanagements. Analysen der Energiedaten und Lastspitzen zeigen auf, dass der Energieverbrauch während des Vorlesungsbetriebs höher ist als zu vorlesungsfreien Zeiten. Bei der Untersuchung von regelmäßig auftretenden Lastspitzen wurden Bereiche für ein automatisiertes Leistungsmanagement identifiziert. Es wurde jedoch auch ersichtlich, dass die Infrastruktur der Messgeräte für weitere Analysen erweitert werden muss. Auf der Grundlage einer erweiterten Infrastruktur, können Forschungen in den Bereichen automatisiertes Leistungsmanagement, Anomalieerkennung technischer Geräte und Elektromobilität, insbesondere hinsichtlich der Nutzung erneuerbarer Energien in einem Smart Grid, erfolgen.

Keywords: Smart Grid, Energie- und Leistungsmanagement, Data Monitoring

1 Einleitung

Vor dem Hintergrund der Minimierung ökologischer und gesellschaftlicher Probleme im Zuge der Energiewende ist neben dem Übergang von fossilen Energieträgern und der Kernenergie hin zu erneuerbaren Energien insbesondere auch der Bedarf an intelligenten Stromnetzen, sogenannten Smart Grids, gestiegen [Le15]. In solchen Netzen muss die Optimierung und Überwachung der Energielandschaft durch die kommunikative Vernetzung und Steuerung von Netzbetriebsmitteln (Stromerzeugern/-verbrauchern) in Energieübertragungs- und verteilungsnetze ermöglicht werden [In13].

Durch die Aufstellung relevanter Konzepte und Komponenten, die dazu geeignet sind regulierend in den Energiehaushaltsplan einzugreifen, kann das Potential eines Smart Grids verbessert werden [CWS07]. Abweichend von der bisherigen nachfrageorientierten Versorgung sind in der heutigen Zeit eher Mechanismen gefragt, die den Bedarf an das Angebot der Einspeisung erneuerbarer Energien in diese Netze angleichen. Die Anforderung der Bildung eines effizienten, dezentralisierten Energiesystems durch Integration von In-

[1] OTH Regensburg, Fakultät IM ,Universitätsstr. 31, 93053 Regensburg, susanne.kenner@st.oth-regensburg.de
[2] OTH Regensburg, Fakultät IM, Universitätsstr. 31, 93053 Regensburg, raphael.thaler@st.oth-regensburg.de

formations- und Kommunikationstechnologien muss durch die Verschiebung gerade nicht benötigter Energievorräte nachgekommen werden. Durch die Verschiebung elektrischer Lasten kann die Spitzenlast reduziert werden und somit ein wichtiger Beitrag zur Reduktion von CO_2-Emissionen und der Erfüllung der Klimaschutzziele geleistet werden. Von zentraler Bedeutung ist daher die Erkennung solcher „verschiebbaren“ Lasten oder Lastspitzen in einer Energielandschaft. Durch Analysen der bestehenden Datenbasis lässt sich ihr Energiebedarf speziell ermitteln und diskreten Zeiträumen zuordnen. Auf Basis der gesammelten Ergebnisse können Regelalgorithmen für ein automatisiertes Leistungsmanagement entwickelt werden, um die identifizierten Lastspitzen auszugleichen [Pa13].

In diesem Beitrag beschreiben wir ein System zur Datenakquirierung und -speicherung und dessen Architektur am Beispiel des Smart Grids der OTH Regensburg. Ziel des Projektes ist es eine einfache Analysemöglichkeit bereit zu stellen, um zukünftig Lastspitzen zu identifizieren und diese durch ein Leistungsmanagement auszugleichen. In Abschnitt 2 wird auf die Hintergründe der Entstehung des Projektes eingegangen. Anschließend werden in Abschnitt 3 verwandte Projekte aufgeführt. Abschnitt 4 beschreibt die entstandene Softwarelösung, Architektur, Funktionalität und Analyseergebnisse. Abschnitt 5 evaluiert Ergebnisse und es folgt ein Ausblick auf zukünftige Forschungsschwerpunkte.

2 Motivation

Mit dem Ziel der kommunikativen Vernetzung und der Steuerung von Betriebsmitteln forscht die OTH Regensburg seit einigen Jahren im Bereich Smart Grid. In diesem Rahmen wurden u.a. Mobilfunktechnologien auf ihre Eignung zur Datenakquirierung im Hinblick auf Übertragungsrate, Latenz, Sicherheit und Verlässlichkeit getestet [Ze13], [Sc14]. Die OTH Regensburg bedient sich für ihre Forschung eines breiten Spektrums an Messgeräten, die campusweit in das Stromnetz integriert sind. Im Regensburger Center of Energy and Resources (RCER), Forschungscluster Robotik, Algorithmen, Kommunikationsnetze und Smart Grid (RAKS) und dem IT- Anwenderzentrum sind unter anderem die Elektromobilität, Energieeffizienz und Energie- und Kommunikationsnetze bezüglich Smart Grid von besonderem Interesse. Zusammen mit der Fakultät Elektro- und Informationstechnik (EI) wird ein Forschungsprojekt zum Leistungsmanagement des Stromnetzes der OTH Regensburg durchgeführt. In diesem Projekt umfasst unser Arbeitspaket die Kommunikation, Bereitstellung und Visualisierung der Daten des Smart Grids.

Um eine Übersicht der Energielandschaft der Hochschule Regensburg zu generieren, wurde der sogenannte „Smart Energy Campus“ ins Leben gerufen, der sich die bestehende Gebäudeinfrastruktur zu Nutze macht und so eine hochauflösende Transparenz bietet. Hierbei handelt es sich um eine von Studenten kontinuierlich weiterentwickelte Softwarelösung in Java, die Messdaten vieler Messgeräte (im Folgenden auch Smart Meter genannt) über eine Modbus-TCP Verbindung bis zu einer sekündlichen Auflösung erfasst und in einer Apache Cassandra Datenbank speichert. Auf dieser bereits vorhandenen Softwarlösung wurde das System in diesem Projekt um Funktionaliäten erweitert und um eine Analysesoftware ergänzt. Zudem wurde eine Geräte- und Benutzerverwaltung aufgesetzt, die die Administration der eingesetzten Geräte gestattet. Die Analyse der Daten wird durch

eine Weboberfläche realisiert, die zum einen eine Topologieansicht der Hochschule bietet und zum anderen aber auch ausgewählte Daten einzelner Messstationen in diversen Diagrammen darstellen kann. Somit wird es Studenten ermöglicht sich bei ihren Projekten und dem Übungsbetrieb auf tatsächliche Daten zu beziehen. Von besonderem Interesse für diese Parteien ist die sekündliche Skalierung der Messdaten, da gerade diese ausschlaggebend für ihre Forschung sind. Die Projektbeteiligten der Fakultät EI analysieren das Smart Grid und ermitteln Leistungsmanagementpotenziale. Durch die vorgestellte Software ist es möglich die Daten des Stromnetzes zu analysieren und Lastspitzen zu identifizieren.

3 Verwandte Projekte

Erste Schritte bezüglich leistungsorientierter Eingriffe in ein Smart Grid wurden bereits durch die Einführung regulierender Systeme in kleineren Testenergielandschaften durchgeführt [Ba08]. Ebenso wurden Systeme getestet, die unter Einbezug der erneuerbaren Energien versuchen, Energie genau dort im verteilten System zugänglich zu machen wo diese benötigt wird [CWS07]. Um eine vorübergehende Speicherung und Verteilung der Energie zu gewährleisten kommen Batterien und Akkumulatoren zum Einsatz [Pa13], [Zh11]. Ein Beispiel wie zukünftig mittels intelligenter Ladeinfrastruktur und innovativen Steuerungs- und Kommunikationskonzepten im Bereich Elektromobilität Energie gespart werden kann stellt das Projekt E-Wald dar [Sp15].

Mit der Netzvisualiserungssoftware GridVis wird aktuell von Seiten des Technischen Betriebs (TB) auf die Messgeräte zugegriffen. Frei definierbare Topologieansichten erlauben eine Übersicht der Verteilung des elektrischen Netzes, wobei die zeitliche Synchronisation aller Geräte über einen NTP-Server erfolgt. Die Benutzerverwaltung regelt mittels definierbarer Projekte und Rollen die erlaubten Zugriffe involvierter Parteien. Generell ist die Software für Janitza-Geräte optimiert, für die alle Geräte bereits vollständig parametriert adressiert werden können. Fremdgeräte können als virtuelle Geräte über das *generic Modbus Protokoll* eingebunden werden. Bei derartigen Geräten können lediglich aggregierte Minutenwerte gespeichert werden. Somit entfallen die für die Forschung interessanten Sekundenwerte. GridVis speichert die Werte abhängig von der erworbenen Lizenz in einer relationalen Datenbank. Zur Anzeige von Daten muss zwischen Online-Daten (Momentanwerte) und historischen Daten (aufgezeichnet) unterschieden werden. Erfasste Daten können über die REST-Schnittstelle (Representational state transfer) abgefragt werden [Br14].

4 Energy Campus OTH Regensburg

Das Smart Grid der OTH besteht derzeit aus ca. 35 Smart Metern, die Daten aus dem Stromnetz zur Verfügung stellen. Neben den Haupttransformatoren und markanten Punkten mit großen Verbrauchern wird auch die erzeugte Energie einer Photovoltaikanlage erfasst. Zu Forschungszwecken erfolgt das Auslesen der Messgeräte sowie die Darstellung der Daten mittels eigener Implementierungen. Die ursprüngliche Architektur zeigt Abb. 1 und wird in Abschnitt 4.1 detaillierter erläutert. In Abschnitt 4.2 wird der während des

Projekts durchgeführte Architekturumbau und die zukünftige Zielarchitektur beschrieben. Mittels einer Weboberfläche (Abschnitt 4.3) wurden Analysen des Smart Grids durchgeführt (Abschnitt 4.4).

4.1 Ausgangsarchitektur

Hardware: Abb. 1 zeigt die in drei Hauptbreiche aufgeteilte Gesamtarchitektur des Smart Energy Campus. Jeder Bereich wird durch einen virtuellen Server mit Windows Server Betriebssystem repräsentiert. In das Stromnetz der OTH Regensburg wurden bereits in früheren Projekten, in Zusammenarbeit mit dem TB, Smart Meter der Firmen Siemens (Siemens PAC4200) [Si10] und Janitza (UMG96RM) [Ja15] eingebaut und über Ethernet mit dem IT-Netzwerk der Hochschule verbunden. Beide Gerätetypen bieten Modbus over TCP/IP als Kommunikationsprotokoll an, über welches verschiedene Messdaten erfasst und gespeichert werden. Das kleinste Ausleseintervall beträgt dabei 200 ms.

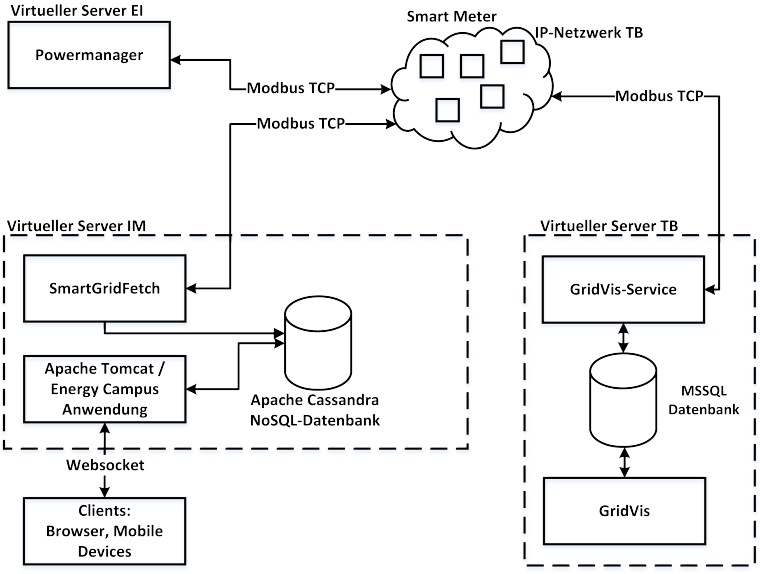


Abb. 1: Ausgangsarchitektur

Software: Wie in Abb. 1 ersichtlich, arbeiten die Fakultät EI und der TB mit kommerziellen Produkten. Die Applikation Siemens-Powermanager wird in der Fakultät EI im Vorlesungsbetrieb genutzt, während der TB auf die Software GridVis der Firma Janitza setzt. Beide Applikationen greifen über Modbus TCP/IP auf die Smart Meter zu und speichern die erfassten Daten in relationalen Datenbanken. Der Siemens-Powermanager bietet viele Möglichkeiten zur Visualisierung der Daten, jedoch zeigen sich Schwächen bei dem Auslesen von Messgeräten anderer Hersteller und bezüglich des Exports von Daten. Für den TB ist insbesondere die Möglichkeit der Generierung von Kostenreports von Bedeutung. Durch die bereitgestellte Topologieansicht bietet GridVis einen guten Überblick über die Messgerätelandschaft und dient der Kostenkontrolle. Beide Systeme können sekündliche

Messwerte auslesen, diese aber nicht in den Datenbanken speichern, was jedoch gerade zu Forschungszwecken interessant ist.

Um größtmögliche Unabhängigkeit von kommerziellen Softwarelösungen zu garantieren und Schwächen dieser auszugleichen wurden eigene Lösungen implementiert, die auf einen Forschungsbetrieb in einem Smart Grid ausgerichtet sind. Auf einem Server der Fakultät Informatik und Mathematik (IM) kommt die Eigenimplementierung zum Einsatz. Diese besteht im Wesentlichen aus der Java-Applikation SmartGridFetch, welche Messdaten über das Modbus Protokoll ausliest und in einer Apache Cassandra Datenbank speichert. Dieses Erfassungssystem wurde bereits in einem früheren Projekt umgesetzt. Bei der Cassandra Datenbank handelt es sich um eine nicht relationale NoSQL (not only SQL) Datenbank. Cassandra bietet ein flexibles Datenbankschema, welches sich bei der Änderung der Infrastruktur beliebig anpassen lässt. Cassandra bietet darüber hinaus einfache Mechanismen zur horizontalen Skalierbarkeit. Im Gegensatz zu den meisten relationalen Datenbanken, welche vertikal skalieren, kann bei ansteigender Datenmenge die NoSQL Datenbank beliebig durch hinzufgen weiterer Serverknoten erweitert werden. Dies ist in einem Smart Grid aufgrund des hohen Datenaufkommens von Vorteil [Ar10], [NPP13]. Zur Visualisierung und Analyse der Messdaten dient die Webapplikation Energy Campus, welche in 4.3 näher erläutert wird.

4.2 Zielarchitektur

Aufgrund der wachsenden Anzahl an Projekten und Applikationen im Smart Grid der OTH Regensburg kommt es u.a. zu Zugriffskonflikten auf die Smart Meter. Tests zeigten, dass die Siemens PAC4200 maximal zwei gleichzeitige TCP/IP-Verbindungen eingehen können. Die zu Beginn des Projekts geplanten Gerätezugriffe waren schnell ausgelastet, wodurch sich Datenverluste in allen Systemen zeigten. Dieser Umstand und IT-Sicherheitsrelevante Überlegungen führten zu einem Architekturumbau. Abb. 2 zeigt die geplante Zielarchitektur, die größtenteils bereits umgesetzt wurde. Die Zugriffe auf die Smart Meter werden zukünftig durch eine zentrale Ausleseinstanz entkoppelt. Aus Sicherheitsgründen befinden sich alle Geräte in einem abgeschotteten IP-Netz des TB. Nur die Software GridVis des TB hat Zugriff auf die Smart Meter. Somit dient GridVis als zentrale Datensammelstelle, über deren REST-Schnittstelle die abgerufenen Daten weiteren Projektbeteiligten und Applikationen zur Verfügung gestellt werden. Die Applikation GridVisFetch ruft die Daten sekündlich und in einem Intervall von 15 Minuten ab und speichert diese in der bestehenden Apache Cassandra Datenbank. Von dort aus werden weitere Systeme bedient. Im Hinblick auf ein Leistungsmanagement und Echtzeitmonitoring sind sekündliche Abfragen von Bedeutung. Hier ist noch zu prüfen, ob dies bei einer größeren Messgerätelandschaft effizient über die REST-Schnittstelle erfolgen kann. Um Mehrfachzugriffen auf die Smart Meter vorzubeugen soll der Powermanager in Zukunft über virtuelle Geräte versorgt werden. Dies wurde jedoch noch nicht im Rahmen dieses Projektes umgesetzt. Aufgrund der Netzabschottung der Smart Meter, wurde eine Testumgebung aufgesetzt, in der ebenfalls ins Stromnetz integrierte Geräte verschiedener Hersteller zu Forschungszwecken zur Verfügung stehen. Während die Hoheit und Administration der Geräte beim TB liegt, können dennoch weitere Beteiligte das Smart Grid der

OTH Regensburg erforschen und analysieren. Ein zusätzlicher Vorteil der Zielarchitektur ergibt sich aus den bereits in der Software GridVis eingebundenen Wasser- und Heizungsmessgeräten, deren Werte ebenfalls über die REST-Schnittstelle zur Verfügung gestellt werden können.

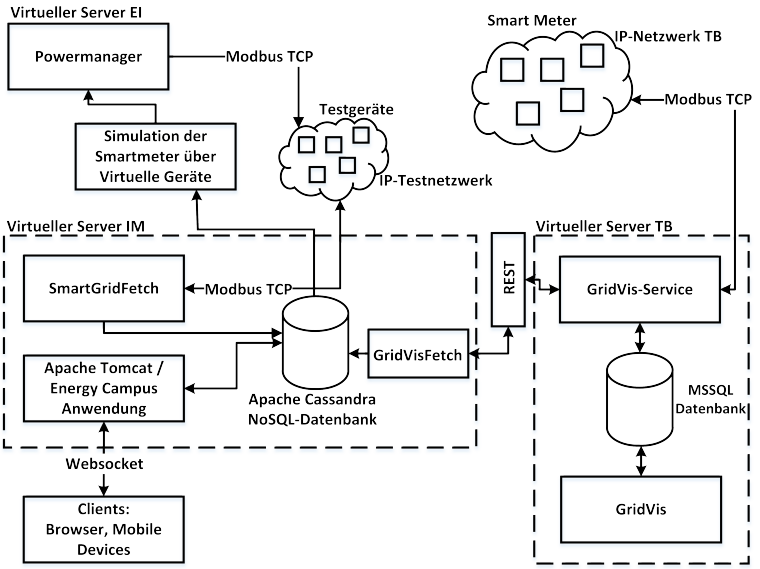


Abb. 2: Zielarchitektur

4.3 Weboberfläche

Die Weboberfläche dient der Darstellung und Analyse der erfassten Messdaten und wurde in HTML5 und JavaScript realisiert. Somit ist ein plattformunabhängiger Zugang verschiedener Clients möglich. Der Client kommuniziert über Websockets mit einer Serverapplikation, die auf einem Apache Tomcat 8 Applikationsserver als Middleware zum Einsatz kommt. Die Serverapplikation bildet die Persistenzschicht auf die Datenbank ab und führt notwendige Berechnungen durch. Mittels Websockets ist eine sessionbehaftete bidirektionale Kommunikation zwischen Client und Server möglich [PN12], [We11]. In unserem Projekt wird dies zur Darstellung von Echtzeitdaten und Senden von Push-Notifications (Warnmeldungen) genutzt. Im Hinblick auf ein automatisiertes Eingreifen in das Stromnetz bietet diese Technologie aufgrund der geringen Latenzzeiten zusätzlich Vorteile.

Die Weboberfläche besteht aus den vier Bereichen Dashboard (Abb.3), Analyse, Map und Geräte- und Nutzerverwaltung. Auf dem Dashboard werden Echtzeitdaten zum Stromverbrauch sowie zur Stromerzeugung angezeigt. Im Analysebereich ermöglichen auf die Forschung ausgerichtete Diagramme die Identifizierung von Lastspitzen sowie deren zeitliche Lage. Es können sowohl die Verbrauchsdaten einzelner als auch der direkte Vergleich der Daten mehrerer Messgeräte dargestellt werden. Diagramme bezüglich der Datendichte

einzelner Messgeräte können generiert werden. Die Datendichte beschreibt den Prozentsatz der Menge der tatsächlich vorhandenen gegenüber der erwarteten Daten innerhalb eines Zeitraums. Aufgrund der oben erwähnten Zugriffskonflikte und gelegentlichen Wartungsarbeiten an den Servern, kommt es dazu, dass Daten nicht kontinuierlich abgerufen werden können. Durch die Datendichte, kann der Prozentsatz an vorhandenen Daten Aufschluss über die Aussagekraft der Analysen geben. Die Daten aller Diagramme können ins CSV-Format (Comma-separated values) exportiert und in entsprechenden Anwendungen weiterverarbeitet werden. Hierdurch kann der Nutzer noch fehlende Analysefunktionalitäten ausgleichen und die Daten frei bearbeiten. Dies bietet dem Nutzer einen Vorteil, da er im Gegensatz zu kommerziellen Produkten nicht von der Software abhängig und auf deren Funktionalität beschränkt ist. Während der Analysebereich eine Sicht auf die Smart

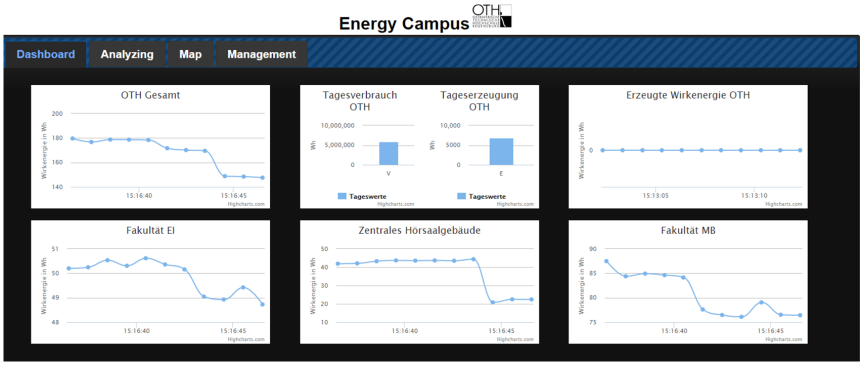


Abb. 3: Dashboard

Meter laut Netzplan gewährt, zeigt der Map-Bereich eine weitere Topologieansicht, in der alle Smart Meter bezüglich ihres Messstandorts in einer Kartenansicht dargestellt werden. Hier können unterschiedliche Messwerte in einem Diagramm verglichen werden, um z.B. Abhängigkeiten der Messwerte untereinander zu erkennen. Der letzte Bereich stellt eine einfache Geräte- und Nutzerverwaltung dar.

4.4 Analyseergebnisse

Bereits Mitte 2013 begann die Datenerfassung des Smart Grids prototypisch. Durch Softwarefehler und Zugriffskonflikte der unterschiedlichen Forschungsbeteiligten schwankte die Datendichte der erfassten Daten erheblich. Erst durch den kontinuierlichen Ausbau der Software und dem Beseitigen der Zugriffskonflikte kann durchgängig eine Datendichte von 98-100% erreicht werden. Die geringen Datenverluste entstehen durch Wartungsarbeiten an den Servern, die meist zu Nachtzeiten erfolgen. Um Aussagen über Einsparpotentiale hinsichtlich eines Leistungsmanagements zu treffen sind umfangreiche Analysen der derzeitigen Situation im Stromnetz der OTH Regensburg nötig. Folgende Punkte beschreiben Analysen und geben eine Übersicht über die Funktionalität der implementierten Software. Aus datenschutzrechtlichen Gründen wurden die Absolutwerte der Analysedaten verändert bzw. anonymisiert, das Verhältnis bei Vergleichen wurde jedoch beibehalten.

Verbrauchsanalyse: Für einen allgemeinen Überblick wurde zunächst der Energieverbrauch der OTH Regensburg evaluiert. Ausgehend von den Haupttransformatoren der verschiedenen Gebäude haben wir die Analyse auf einzelne Messgeräte ausgeweitet. Es zeigt sich, dass der Bereich EI mit 46% den prozentual größten Verbraucher darstellt. Gefolgt von Maschinenbau (MB) mit 33% und dem Zentralen Hörsaalgebäude (ZG) mit 21%, in dem die Mensa und Bibliothek erfasst werden. In den Bereichen EI und MB befinden sich verhältnismäßig viele technische Labore sowie ein Reinraumlabor in dem die Temperatur und Luftfeuchtigkeit konstant gehalten werden muss, wodurch der Stromverbrauch sehr hoch ist. Bei den Analysen ist wie erwartet zu beobachten, dass der Gesamtverbrauch der

Bereich	Vorlesungsfreie Zeit	Vorlesungsbetrieb
EI	69,4%	100%
MB	69,9%	100%
ZG	56,8%	100%

Tab. 1: Vergleich des Stromverbrauchs zwischen vorlesungsfreier Zeit und Vorlesungsbetrieb in %

OTH Regensburg in den vorlesungsfreien Monaten (August-September 2014) deutlich geringer ist als während des Vorlesungsbetriebs (Oktober-November 2014). Tabelle 1 zeigt den prozentualen Unterschied der drei Bereiche EI, MB und ZG. Im Bereich ZG ist der stärkste Anstieg zu Beobachten. Hier ist ersichtlich, dass mit Semesterbeginn die vermehrte Nutzung von Bibliothek und Mensa eine große Auswirkung auf den Verbrauch hat. In den Bereichen EI und MB steigt der Verbrauch ebenfalls, jedoch im Vergleich zum Bereich ZG etwas geringer, an. Dies lässt darauf schließen, dass die große Anzahl an technischen Einrichtungen sowie das Reinraumlabor eine hohe Grundlast erzeugen.

Lastspitzenanalyse: Automatisierte Lastspitzensenkung ist ein zentrales Thema dieses Projekts. Um dies zu realisieren müssen zunächst die Lastspitzen und mögliche Regelmäßigkeiten im Smart Grid der OTH identifiziert werden. Ausgehend von den größten Posten der Verbrauchsanalyse wurden hierarchisch weitere Geräte untersucht. Dabei wurden der Betrag der Wirkenergie sowie das zeitliche Auftreten der Lastspitzen analysiert.

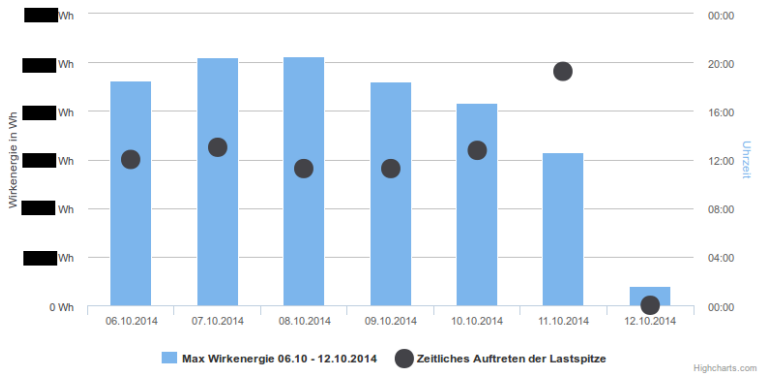


Abb. 4: Zeitliche Lage der Lastspitzen der Mensa

Abb. 4 zeigt einen Ausschnitt aus der Analyse der Mensadaten. Das Diagramm veranschaulicht die zeitliche Lage der täglich aufgetretenen maximalen Lastspitzen einer Woche

im Oktober 2014. Der Betrag der Lastspitzen ist als Balkendiagramm auf der Primärachse und deren zeitliche Lage auf der Sekundärachse dargestellt. Es ist ersichtlich, dass die Lastspitzen meist zur Mittagszeit vorkommen. Deutliche Abweichungen zeichnen sich an Wochenendtagen ab, an denen wenig bis gar kein Betrieb herrscht und somit die Lastspitze von der Grundlast der kontinuierlichen Verbraucher abhängt. Bei Lastspitzenanalysen von weiteren Bereichen bzw. Messgeräten waren keine derart eindeutigen Regelmäßigkeiten im Bezug auf deren zeitliches Auftreten erkennbar. Die Infrastruktur der Smart Meter muss weiter ausgebaut werden, um detailliertere Analysen durchzuführen und nachfolgende Projekte voranzutreiben. Die folgenden Analysen zeigen Beispiele, welche Möglichkeiten die Software bietet. Nach einer Erweiterung der Messinfrastruktur können solche Analysen Aufschluss für ein Leistungsmanagement geben.

Analysen am Beispiel der Stromdaten von Aufzügen: Im Folgenden werden Analysen am Beispiel der Stromdaten von Aufzügen dargestellt. Abb. 5 zeigt eine typische Woche zur vorlesungsfreien Zeit und Abb. 6 während des Vorlesungsbetriebs. Die Diagramme stellen die zu 15-Minutenwerten gemittelten Energiedaten in kWh dar. Im Vergleich beider Diagramme ist ersichtlich, dass zur vorlesungsfreien Zeit der Datenverlauf sowie die Lastspitzen unregelmäßiger auftreten. Während des Vorlesungsbetriebs treten die Lastspitzen zu Beginn des Tages auf. Vergleicht man die Daten mit den Stundenplänen des untersuchten Gebäudes, so steigt der Energieverbrauch der Aufzüge bis auf wenige Ausnahmen proportional zu den gehaltenen Vorlesungen. Lastspitzensenkung könnte hier durch die zeitliche Verlegung von Vorlesungen (Stichpunkt: „energieeffiziente Stundenpläne“) oder durch die Verlegung der Vorlesungen in das Erdgeschoss erreicht werden.

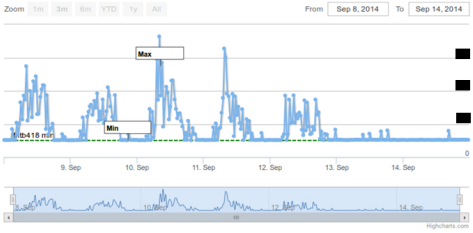


Abb. 5: Lastspitzen Aufzüge vorlesungsfreie Zeit

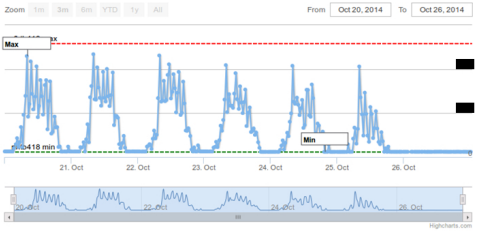


Abb. 6: Lastspitzen Aufzüge Vorlesungszeit

Analyse der erzeugten Solarenergie: Aussagen über die Solaranlage des Smart Grids der OTH Regensburg können ebenfalls getroffen werden. Abb. 7 zeigt für die Monate August 2014 bis Oktober 2014 den Betrag der erzeugten Solarenergie in kWh und vergleicht diese mit den erfassten Sonnenstunden des Deutschen Wetterdienstes, die auf der rechten Sekundärachse eingetragen sind. Aus dem Diagramm geht hervor, dass sich die erzeugte Energie in etwa proportional zu den Sonnenstunden verhält.

An der OTH Regensburg gibt es Forschungen im Bereich der Elektromobilität. Zu diesem Zweck wurde eine Ladestation für Elektrofahrzeuge in die Energielandschaft des Smart Grids integriert. Mit Hilfe der erfassten Stromdaten können Analysen zu den Projekten in diesem Bereich erfolgen. Abb. 8 zeigt ein Beispiel, wie die Analysesoftware zu diesem Zweck genutzt werden kann. Diese zeigt eine Gegenüberstellung des Verbrauchs der

Ladestation und der erzeugten Energie der Solaranlage in kWh. Integriert man die erzeugte Solarenergie in ein Leistungsmanagement so können in Zukunft anhand solcher Diagramme die entwickelten Algorithmen zur effizienten Nutzung der Solarenergie überprüft werden.

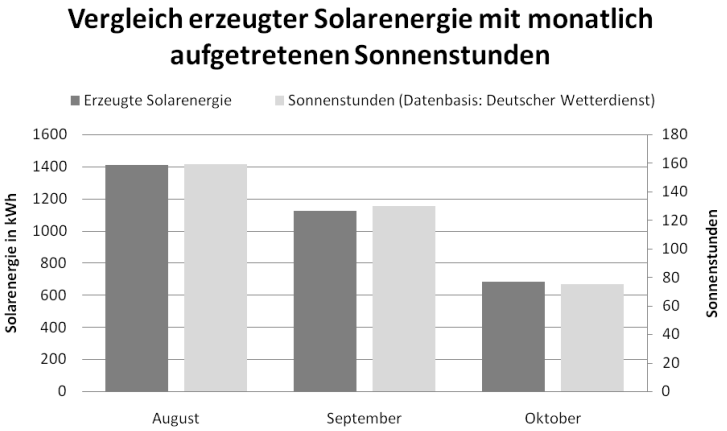


Abb. 7: Vergleich der erzeugten Solarenergie mit den monatlich aufgetretenen Sonnenstunden

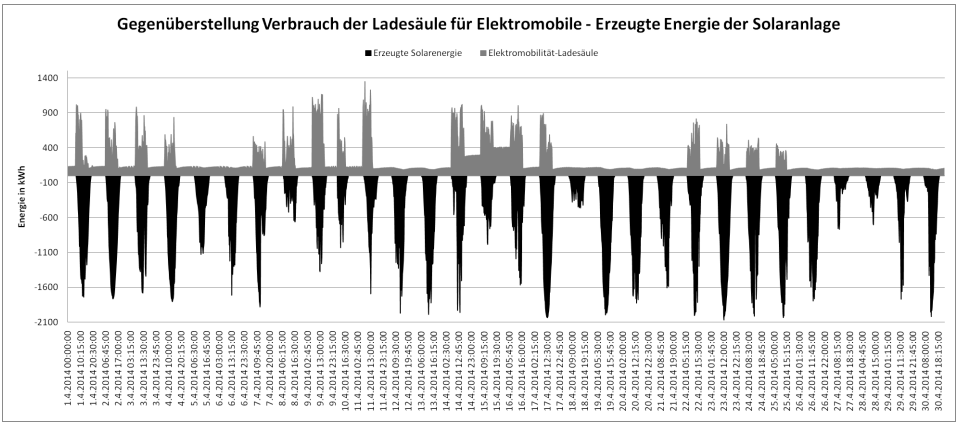


Abb. 8: Analyse der erzeugten Solarenergie und des Verbrauchs einer Ladesäule für Elektromobile

5 Fazit und Ausblick

In diesem Beitrag wurde die Smart Grid Basisarchitektur der OTH Regensburg sowie die Zielarchitektur vorgestellt. Zur Erfassung, Visualisierung und Analyse der Netzdaten wurde eine Software implementiert und in die Systemarchitektur integriert. Die vorgestellte Applikation bildet die Grundlage zur Analyse der Energielandschaft der OTH Regensburg. Die Analysen in Abschnitt 4.4 zeigen einen Ausschnitt aus dem Smart Grid der

OTH Regensburg und verschaffen einen Überblick über den Energieverbrauch. Aus den Analysen der Lastspitzen geht hervor, dass in wenigen Bereichen Regelmäßigkeiten zu erkennen sind. Auf der Grundlage dieser regelmäßig auftretenden Lastspitzen müssen Algorithmen zur automatisierten Lastspitzensenkung erarbeitet werden, um den Energiebedarf zu senken. Abb. 8 stellt ein Beispiel dar, wie erneuerbare Energien zum Laden von Elektrofahrzeugen genutzt werden können. Weitere Forschungen im Bereich Elektromobilität und das Einbinden von volatilen Energieerzeugern in ein bestehendes Stromnetz werden durch die Analysesoftware erleichtert. Ein weiterer Forschungsansatz ergibt sich exemplarisch aus den Analysen der Energieverbrauchswerte von Aufzügen. Aus den Daten kann die Benutzungshäufigkeit bestimmt werden. Somit lässt sich unter der Betrachtung viertelstündlicher Werte beispielsweise der Zeitpunkt von Vorlesungen erkennen. Durch die Senkung des Intervalls auf Minuten- oder sogar Sekundenwerte wären genauere Analysen zur aktuellen Traglast möglich. Die Betrachtung der Spannungswerte einzelner Verbraucher könnte Aufschluss über das Verhalten bei auftretenden Anomalien geben. Hierdurch ist ein Eingreifen vor Eintreten von technischen Defekten möglich.

Zwar erlaubt die Softwarelösung die Erkennung und Einordnung von Lastspitzen zu den jeweiligen Messstellen, um jedoch automatisiert regulierend in das System einzugreifen, muss die Infrastruktur der Messgeräte erweitert werden. Derzeit ist die Granularität der vorhandenen Smart Meter für dieses Szenario zu gering. Unter der Voraussetzung eines detaillierten Netzplans und der zusätzlich eingesetzten Messgeräte könnte man beispielsweise über WLAN-Switches oder Raspberry Pis nicht benötigte Verbraucher automatisch zu identifizierten Zeiten vom Stromnetz trennen und somit Energie sparen. Die Software könnte demnach um Regelalgorithmen zur Fernsteuerung von Verbrauchern erweitert werden. Im weiteren Projektverlauf muss noch geprüft werden, ob die Architektur aus Abschnitt 2 für Echtzeitmonitoring und ein Leistungsmanagement im Hinblick auf die Performanz der REST-Schnittstelle in einer größeren Messgerätelandschaft geeignet ist.

Danksagung

Die in dieser Arbeit dargestellten Ergebnisse sind im Rahmen von Projekten an der OTH Regensburg entstanden und wurden durch das Regensburg Center of Energy and Resources (RCER) und dem Technologie- und Wissenschaftsnetzwerk Oberpfalz (TWO) unterstützt. Wir bedanken uns an dieser Stelle auch bei den Projektleitern und Betreuern unserer Arbeiten, Markus Kucera, Klaus Volbert und Thomas Waas.

Literaturverzeichnis

[Ar10] Arenas-Martinez, M.; Herrero-Lopez, S.; Sanchez, A.; Williams, J. R.; Roth, P.; Hofmann, P.; Zeier, A.: A comparative study of data storage and processing architectures for the smart grid. In: Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on. IEEE, S. 285–290, 2010.

[Ba08] Barklund, E.; Pogaku, N.; Prodanovic, M.; Hernandez-Aramburo, C.; Green, T. C.: Energy management in autonomous microgrid using stability-constrained droop control of inverters. Power Electronics, IEEE Transactions on, 23(5):2346–2352, 2008.

[Br14] GridVis-Documentation+5.0, https://wiki.janitza.de/display/GRIDVIS50EN/GridVis-Documentation+5.0, Zuletzt aufgerufen am: 12.04.2015.

[CWS07] Chakraborty, Sudipta; Weiss, Manoja D; Simoes, Marcelo Godoy: Distributed intelligent energy management system for a single-phase high-frequency AC microgrid. Industrial Electronics, IEEE Transactions on, 54(1):97–109, 2007.

[In13] DKE Roadmap, https://www.dke.de/de/std/aal/documents/nr_e-energy%20smart%20grid_de_version%202.0.pdf, Zuletzt aufgerufen am: 12.04.2015.

[Ja15] Manual: UMG96RM, http://www.janitza.de/umg-96rm-e-downloads.html, Zuletzt aufgerufen am: 12.04.2015.

[Le15] Lee, S.; Lim, H.; Go, W.; Park, H.; Shon, T.: Logical Architecture of HAN-Centric Smartgrid Model. In: Platform Technology and Service (PlatCon), 2015 International Conference on. IEEE, S. 41–42, 2015.

[NPP13] Nayak, A.; Poriya, A.; Poojary, D.: Type of NOSQL Databases and its Comparison with Relational Databases. International Journal of Applied Information Systems, 5(4), 2013.

[Pa13] Palma-Behnke, R.; Benavides, C.; Lanas, F.; Severino, B.; Reyes, L.; Llanos, J.; Sáez, D.: A Microgrid Energy Management System Based on the Rolling Horizon Strategy. IEEE Trans. Smart Grid, 4(2):996–1006, 2013.

[PN12] Pimentel, V.; Nickerson, B. G.: Communicating and displaying real-time data with WebSocket. Internet Computing, IEEE, S. 45–53, 2012.

[Sc14] Schlegl, P.; Robatzek, P.; Kucera, M.; Volbert, K.; Waas, T.: Performance Analysis of Mobile Radio for Automatic Control in Smart Grids. In: Proc. of the Second Intl. Conf. on Advances In Computing, Communication and Information Technology-CCIT 2014. Institute of Research Engineers and Doctors, S. 135–141, 2014.

[Si10] Systemhandbuch: SENTRON PAC4200, https://cache.industry.siemens.com/dl/files/595/34261595/att_90266/v1/manual_sentron_pac4200_03_de-DE.pdf, Zuletzt aufgerufen am: 12.04.2015.

[Sp15] E-Wald, https://www.th-deg-de/de./forschung/projekte/e-wald, Zuletzt aufgerufen am: 12.04.2015.

[We11] Wessels, A.; Purvis, M.; Jackson, J.; Rahman, S.: Remote data visualization through websockets. In: Information Technology: New Generations (ITNG), 2011 Eighth International Conference on. IEEE, S. 1050–1051, 2011.

[Ze13] Zenger, G.; Kenner, S.; Volbert, K.; Waas, T.; Kucera, M.: Acquiring energy data from a medium-voltage grid for future smart grid solutions:A practical Smart Grid Application Example realized by use of Cellular Communication Networks of the 2nd and 3rd Generation. In: Intelligent Solutions in Embedded Systems (WISES), 2013 Proceedings of the 11th Workshop on. IEEE, S. 1–8, 2013.

[Zh11] Zhou, H.; Bhattacharya, T.; Tran, D.; Siew, T.; Khambadkone, A. M.: Composite energy storage system involving battery and ultracapacitor with dynamic energy management in microgrid applications. Power Electronics, IEEE Transactions on, 26(3):923–930, 2011.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Building A State-Of-The-Art Model Checker

Sebastian Wolff[1]

Abstract: Verification is the activity of proving a software artefact correct with respect to its specification. In this paper we focus on the algorithmic approach to verification. Therefore, we present how one can build a state-of-the-art model checker for recursive integer programs. Since those programs are present in different environments, many techniques evolved. Inspired by the driver verification with Microsoft's SLAM toolkit, we show how to integrate some of the available techniques into a single tool. One of those techniques is predicate abstraction. It allows us to handle one of today's core problems of model checking, namely a infinite data domain which is introduced by integer variables. Moreover, we integrate a reachability analysis to check the abstraction for correctness. This analysis uses procedure summaries to cope with recursive programs and potentially infinite call stacks. To complete our tool we also integrate a refinement for the abstraction based on Craig interpolation. Altogether, this stack of state-of-the-art techniques allows us to perform a CEGAR loop.

Keywords: Verification, Model Checking, Recursive Programs, Infinite Data Domain.

1 Introduction

Verification is a discipline of computer science which is dedicated to prove software artefacts correct. The question for correct software systems is important as those systems are widespread and deployed in a variety of electronic devices. Among those devices are lots of safety-critical ones, like controller units in cars and aircraft. Quite naturally, we want those systems to operate properly, that is, they should precisely meet their specification. As testing is expensive and not capable of proving correctness, we want to establish techniques that are able to automatically prove systems correct.

Model Checking is an algorithmic verification technique which meets our requirements from above. As it has been around for over 30 years, a lot of different approaches and procedures have been developed for a broad spectrum of systems and software. In this paper, we focus on the class of recursive programs with integer variables.

Recursive integer programs are, however, challenging as they are complex in two dimensions. On the one hand, they feature unbounded-depth recursion. On the other hand, they allow the usage of integer variables which have a potentially infinite domain. Even when restricting integers to 64 bit, as in modern processors, the search space is too big to be explored entirely [He04]. One widespread approach to tackle those challenges is the so called *Counter-Example Guided Abstraction Refinement (CEGAR)* loop [Cl00] depicted in Figure 1. The main idea is to discard unimportant aspects yielding a smaller abstract system. This abstract system is then steadily analysed and refined until it may be proven

[1] Technische Universität Kaiserslautern, s_wolff09@cs.uni-kl.de

correct or incorrect. Unfortunately, termination of the CEGAR approach is not guaranteed due to recursive programs with integer variables being Turing complete [Mi67].

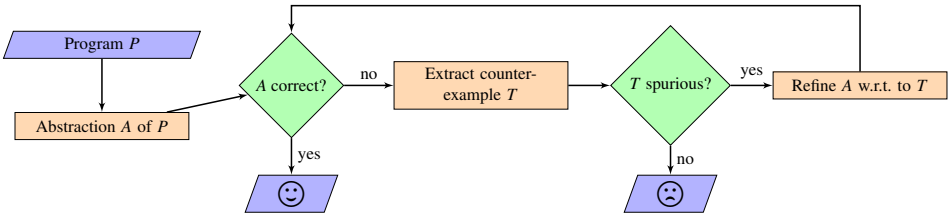


Fig. 1: The CEGAR loop.

A promising approach for generating a smaller system is *Predicate Abstraction*. Its working principle is to use a set of predicates to describe a subset of a system's behaviours. It provides multiple benefits for a model checker. Firstly, if conducted properly, it allows reasoning about the original program via the abstraction [JM09]. Secondly, the abstraction has a boolean, hence finite, data domain tackling one dimension of complexity.

The remaining dimension of complexity is introduced by recursion. To address this challenge, we employ a technique called *procedure summaries*. This approach computes the impact of calling a procedure on a system's state. By combining this technique with a reachability analysis, we are able to check finite systems, like our abstractions, for correctness.

Inspired by the Microsoft SLAM toolkit for static driver analysis, we implemented the CEGAR loop in our tool *RocketScience*[2]. Our contribution is to present how to integrate available state-of-the-art techniques into a functioning model checker. Therefore, we go along an execution of our tool. The input to our tool is a recursive integer program P as described in Section 2. Then, a boolean abstraction $B(P)$ of the input program is generated via the predicate abstraction from Section 3. This new boolean program $B(P)$ is checked for correctness. Therefore, the program is translated into a control flow graph and a reachability analysis with procedure summaries is conducted as presented in Section 4. As a result of the reachability analysis, the program may be proven correct. If this is not the case, a counter-example T is generated and checked for validity according to the procedure from Section 5. If T reveals to be valid, P is shown to be incorrect. Otherwise, refinement, which is described in Section 6, is issued to remove the spurious counter-example from the abstraction. This completes the CEGAR loop and the procedure is repeated.

2 Recursive Programs

In the following, we give the definition of a simple programming language, which is inspired by curly braces languages, mainly C. For a formal definition consider Listing 1. The main features are recursive function calls, variables local to functions, global variables and integer variables with an unlimited domain. For the sake of simplicity, functions

[2] The source code is available at: `https://github.com/Wolff09/RocketScience`

```
Program    ::= VarDef* FunDef*
VarDef     ::= [int | bool] VarName ;
FunDef     ::= void FunName () { VarDef* Statement* }
Statement  ::= if ( BoolExpr ) { Statement* }
             | if ( BoolExpr ) { Statement* } else { Statement* }
             | while ( BoolExpr ) { Statement* }
             | FunName () ;
             | VarName [, VarName]* = Expr [, Expr]* ;
             | assert( BoolExpr ) ;
             | ; // skip
Expr       ::= BoolExpr | IntExpr | VarName | Literal
Literal    ::= true | false | <integer>
BoolExpr   ::= IntExpr [> | < | >= | <= | == | !=] IntExpr
             | BoolExpr [&& | ||] BoolExpr
             | BoolExpr ? BoolExpr : BoolExpr
             | ! BoolExpr
IntExpr    ::= IntExpr [+ | - | * | /] IntExpr
             | - IntExpr
```

List. 1: Language definition.

do not feature formal parameters and return values. This is no limitation to the expressibility of our language because communication between caller and callee can be handled via global variables. To that end, the caller writes the actual parameters to global variables and the callee copies them into local variables. Return values can be handled in the same way.

For the rest of the paper we assume that programs satisfy the following constrains: (a) expressions and statements are properly typed, (b) no global variable is shadowed by a local variable, (c) no variable appears more than once on the left hand side of an assignment, (d) the first statement of a function initializes all local variables, and (e) there is a `main` function which additionally initializes the global variables. These assumptions can be checked easily as they are of a static kind[3].

A *boolean* program is a regular program which is restricted to variables and literals of type `bool`. Additionally, boolean programs support `assume` statements which coincide with `assert` statements beside the fact that they never fail, i.e. simply block the control flow when their condition evaluates to *false*. Furthermore, we allow boolean programs to be non-deterministic by adding the literal *unknown*. These two additions to boolean programs compared to ordinary integer programs come in handy during the abstraction described in the next section.

3 Abstracting Integer Programs

A crucial part of model checking infinite state systems is an abstraction of the original integer program into a boolean program. A well-known approach is predicate abstraction [GS97, JM09]. It divides the infinite state space of the integer program into finitely many equivalence classes. These classes are characterized by a set of predicates which are assertions about the states of the integer program.

[3] Since our language is statically typed even (a) can be checked statically.

In the following, an abstraction procedure is described which is due to Ball et al. [Ba01]. The input to the abstraction is an integer program P and a set of predicates *Preds*. Each predicate $p \in Preds$ is a first order logic formula and comes with a scope. This scope is either *global* or a function. Naturally, global predicates may only range over global variables of P, whereas local predicates may additionally range over local variables of the corresponding function.

The output of the abstraction procedure is a boolean program, denoted by $B(P)$, which has a boolean variable x_p for every predicate $p \in Preds$. The abstraction $B(P)$ imitates the behaviour of P by updating its boolean variables in such a way that they capture how executing a statement from P changes the truth of the predicates. Therefore, the original control flow is preserved and assignments to x_p, $p \in Preds$, replace the statements from P. An example abstraction can be found in Listing 2.

```
// program ExP
int x;
int y;

void main() {
  x, y = 5, 13;
  swap();
  assert(x > y);
}

void swap() {
  x = x + y;
  y = x - y;
  x = x - y;
}

// predicates (global)
p = x <= 5;
q = y >= 13;
```

```
// abstraction B(ExP) w.r.t. {p, q}
int p;
int q;

void main() {
  p, q = true, true;
  swap();
  if (unknown) { assume(!p || !q); }
  else { assume(true); assert(false); }
}

void swap() {
  p = !p && q ? false
              : unknown; // act1
  q =  p && q ? false
              : unknown; // act2
  p =  p && q ? true
              : unknown; // act3
}
```

List. 2: Example program and its abstraction.

Weakest Preconditions Central to the abstraction are weakest preconditions. The weakest precondition $wp(s,\varphi)$, for some statement s and some first order formula φ, is the weakest predicate the truth of which before s entails the truth of φ after s. The weakest precondition for an assignment $x = e$ is thereby defined as $wp(x = e, \varphi) = \varphi[x \mapsto e]$ where $\varphi[x \mapsto e]$ equals φ with every occurrence of x replaced with e.

We also define a strengthening $F(\varphi)$ of φ. It is the weakest formula implying φ and ranging over the set $\{x_p : p \in Preds\}$. Furthermore, a weakening $G(\varphi)$ will be useful and is defined by $G(\varphi) = \neg F(\neg\varphi)$. We employ both those notions as they take formulas ranging over predicates and produce formulas ranging over the corresponding variables from the abstract program. The intuition behind this is an abstraction from formulas from the integer program to formulas in the abstract program. These basic building blocks of our abstraction procedure are effectively computable and an optimized implementation is given in [Ba01].

Conditionals Given an `if(`c`){...}else{...}` construct from program P, we know at the beginning of the then-branch that c holds. Thus, in the abstract program, we want the then-branch to be executed only if the abstraction does not imply $\neg c$. Remember that we already introduced this notion with $G(c)$ and hence come up with the following rule [Ba01].

```
if (c) {                                  if (unknown) {
  ...                                       assume( G(c) );
} else {          abstraction ⟼           ...
  ...                                     } else {
}                                           assume( G(!c) );
                                            ...
                                          }
```

Note here, that we introduce a non-deterministic choice which is guarded with assume statements. We utilize the non-determinism as both $G(c)$ and $G(\neg c)$ might hold in the abstract program [Ba01]. Additionally, this construct allows to explore both branches in the following reachability analysis. The same approach applies to while loops as follows.

```
while (c) {                               while (unknown) {
  ...             abstraction ⟼             assume( G(c) );
}                                           ...
                                          }
                                          assume( G(!c) );
```

Assignments Consider an assignment $x = e$ in P. This statement may influence the truth of a whole range of predicates, namely those containing x. So the abstraction of the assignment is an assignment again which captures the impact of $x = e$ on all predicates. Therefore, consider a predicate p_i which is modelled by the boolean variable b_i. Then, by definition, p_i is *true* after $x = e$ if $wp(x = e, p_i)$ can be shown to evaluate to *true* under every possible assignment to the free variables. That is, $b_i = true$ is a valid assignment if $F(wp(x = e, p_i))$ holds. Analogously, $b_i = false$ is valid if $F(wp(x = e, \neg p_i))$ holds. However, the predicates might be to weak to prove any assignment valid. Naturally, we assign $b_i = unknown$ in such a case.

According to [Ba01], there is always at most one valid assignment to b. By exploiting this fact, one can come up with the following rule [Ba01].

```
                        b1, ..., bn =
                          F(wp(x=e,p1)) ? true
x = e;   abstraction ⟼                : F(wp(x=e,!p1)) ? false : unknown,
                        ...,
                          F(wp(x=e,pn)) ? true
                                        : F(wp(x=e,!pn)) ? false : unknown;
```

Asserts An `assert(`c`)` is supposed to have no effect if the boolean condition c evaluates to *true* and should fail otherwise. Hence, we handle this statement in the same way as an `if`. We come up with the following rule.

```
                    if (c) {                          if (unknown) {
                      // skip                             assume( G(c) );
assert(c);  translation ⟼  } else {   abstraction ⟼    } else {
                      assert(false);                       assume( G(!c) );
                    }                                      assert(false);
                                                         }
```

Note here, that we introduced an `assert(false)` to identify an assertion error in the abstract program. The abstraction, however, is not recursively continued for the newly introduced assertion.

Calls Function calls are simply copied to the abstraction since they have neither formal parameters nor return values. As functions communicate via global variables, the communication can be observed via global predicates and tracked throughout the entire program. Hence, our approach is sufficient for handling simple function calls.

4 Checking Boolean Programs

The next step in the CEGAR loop, after abstracting the input program, is to check the abstraction for correctness. That is, in our context, to check whether there is an execution of the abstracted program which raises an assertion error. According to our abstraction procedure, assertions in the original program are translated into an `if` construct where only the failing branch contains an `assert(false)`. Thus, we only need to check whether any `assert` statement is reachable. If so, the abstraction is considered incorrect.

The above reachability problem can be solved algorithmically on a graph rather than directly on the program code level. To that end, we first introduce a translation from code to control flow graph and then conduct a reachability analysis on the resulting graph. Note that, although the data domain is finite due to the previous abstraction, the reachability analysis still needs to tackle the challenge of arbitrary large call stacks and possibly non-terminating recursion.

Control Flow Graphs A control flow graph G is a finite directed graph $G = (S, V, T, C)$ with a set of nodes S, a set of boolean variables V, a set of edges T and another dedicated set of edges C which are used for function calls only. The set of variables is split into local and global variables, i.e $V = \mathit{Locals} \uplus \mathit{Globals}$. T-edges additionally come with a *guard* and a set of *actions*. The guard is a first order logic formula and an action has the form $x = e$ with $x \in V$ and a boolean term e denoting the new value of x. For a T-edge from s to s' with guard g and actions $x_1 = e_1, \ldots, x_n = e_n$ we write

$$s \xrightarrow{g/x_1=e_1,\ldots,x_n=e_n}_T s'.$$

A configuration is a tuple $cf = (s, val)$ with state $s \in S$ and $val : V \mapsto \{true, false, *\}$ being a valuation to the variables in V. The valuation val may map some variables to $*$ denoting an arbitrary truth value. If such a mapping exists we call the valuation *partial* and *complete* otherwise. The evaluation of a formula φ based on a valuation val is denoted by $\|\varphi\|_{val}$.

Consider some transition $s \xrightarrow{g/x_1=e_1,\ldots,x_n=e_n}_T s'$ and some configuration $cf = (s, val)$. The transition can be taken by cf if g is enabled, i.e. if $\|g\|_{val} = true$. The result of taking the transition is a new configuration $cf' = (s', val')$ with

$$val'(x) := \begin{cases} \|e_i\|_{val}, & \text{if } x \in \{x_1, \ldots, x_n\} \\ val(x), & \text{otherwise.} \end{cases}$$

From Boolean Programs to Control Flow Graphs Next, we present the translation procedure. Therefore, consider a boolean program $B(P)$ and an empty control flow graph G which is extended as we process $B(P)$ and used as output of our procedure.

Firstly, we generate a skeleton for every function from $B(P)$. That is, for every function we add two nodes to G: an *entry* and an *exit* node both of which are unique. Those skeletons are required to translate recursion properly since a function might invoke itself or another function which has not been (completely) translated yet. Given some function f, we may reference those nodes as *f.entry* and *f.exit*, respectively. Secondly, we add a dedicated *error* node which we use to identify assertion errors. Lastly, we add sufficiently many variables to G such that every variable x from $B(P)$ can be mapped to a variable $x_G \in V$. Hence, an expression e from $B(P)$ can be translated by simply replacing every variable x occurring in e with its corresponding graph variable x_G.

With this basic structure set up, the actual translation of the function bodies from $B(P)$ is straight forward and skipped for brevity. The only statement that needs some special treatment is the function call. A call to function f in $B(P)$ introduces two new nodes in G – the *call* node and the *return* node. Additionally, we add a C edge between the call node and *f.entry* as depicted in Figure 3. Using the dedicated *call* transition relation C allows us to differentiate between sequences of statements and function calls when performing a reachability analysis. This will ultimately allow us to restore local variables in unbounded-depth recursion during the later reachability analysis.

Reachability Analysis with Procedure Summaries Given the control flow graph G resulting from a translation of a boolean program $B(P)$, it remains to conduct a reachability analysis to check whether an assertion error might occur. By construction, we simply need to check whether the dedicated *error* state is reachable in G.

For finite transition systems, the set of reachable configurations can be effectively computed as a fixed point to the equation $x = x \cup post_T(x)$ [Sc04]. This approach is, however, insufficient for our purpose as the post image does not handle local variables properly. Naturally, a function call should not alter the local variables. But if we would simply compute post images for the *call* transitions, we could not restore local variables for returning recursive calls[4]. Hence, we apply a technique called *procedure summaries* [BR00]. It basically executes a separate *sub*-analysis and as a result augments the control flow graph with an additional T-edge which summarises the effect of the function call on the global variables. For an example procedure summary consider Figure 2.

Example path of configurations (states omitted) through the abstract version of `swap()` from Listing 2:

$$\begin{Bmatrix} p \mapsto 1 \\ q \mapsto 1 \end{Bmatrix} \xrightarrow{act1} \begin{Bmatrix} p \mapsto * \\ q \mapsto 1 \end{Bmatrix} \xrightarrow{act2} \begin{Bmatrix} p \mapsto 1 \\ q \mapsto 0 \end{Bmatrix} \xrightarrow{act3} \begin{Bmatrix} p \mapsto * \\ q \mapsto 0 \end{Bmatrix}$$

Resulting summary edge:

$$call \xrightarrow{\substack{p \leftrightarrow 1 \wedge q \leftrightarrow 1 \,/ \\ p = *,\, q = 0}}_T return$$

Fig. 2: Procedure summary example.

[4] Memorizing the local variables of the call site is not possible as the stack of recursive functions might grow beyond all bounds and our control flow graph is required to be finite.

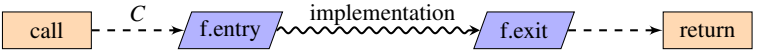


Fig. 3: Call scenario.

A formal description is in order. Therefore, consider a configuration $cf_0 = (call, val_0)$ with respect to Figure 3. Furthermore, assume that *val* is complete. Then, compute the post image of cf_0 relative to the *call* transition relation C and quantify out all local variables. This gives a new (partial) configuration $cf_1 = (f.entry, val_1)$, with $val_1(g) = val_0(g)$ for global variables g and $val_1(l) = *$ for local variables l. Next, conduct a reachability analysis relative to T, as described above, yielding a set of reachable configurations CF. If there is some $cf \in CF$ with $cf = (f.exit, val)$, we can add a *summary edge* to G describing the effect of f to the global variables. Hence, we augment T by adding the edge $call \xrightarrow{g/a}_T return$ with

$$g = \bigwedge_{x \in Globals} x \leftrightarrow val_0(x) \qquad \text{and} \qquad a = \{x = \|x\|_{val} \ : \ x \in Globals\}.$$

Additionally, we might need to recursively invoke this procedure when a new call site is found, i.e. if there is some $cf \in CF$ where the state of cf describes some function's *call* block and no summary has been computed yet. However, when recursively descenting, one must prevent repetitions. That is, when computing the summary for cf, no sub analysis must be issued for cf (again). This is because a reoccurring configuration in the recursive call stack resembles a non-terminating function call in the program. Hence, no summary is computed for such a call. Furthermore, this guarantees termination of our procedure.

Our overall approach for model checking boolean programs simply interleaves both above techniques exhaustively. That is, we compute the fixed point for $x = x \cup post_T(x)$, conduct procedure summaries for all call sites, and repeat this until the set of configurations saturates. Lastly, it remains to check membership of the *error* state.

5 Counter-Example Traces

Applying the procedure described in the previous section might reveal that the abstract boolean program $B(P)$ can run into an assertion error by showing that the dedicated *error* state of the control flow graph associated with $B(P)$ is reachable. In that case, we have to check if the malicious behaviour is also present in the original integer program P. Therefore, we have to compute a so called *counter-example trace* and check whether this trace is valid. A *counter-example trace* is thereby a sequence of statements from P which ultimately runs into an assertion error. Intuitively, it is a linearisation of P which does not contain control structures like *if* and *while* statements. Lastly, checking the validity of the trace means to check whether or not it is a valid execution of P. If so, P is proven incorrect. Otherwise the abstraction $B(P)$ is too imprecise as it allows malicious – so called *spurious* – behaviours which are not present in P.

Generating Traces In the following we give a description of a method for generating a counter-example trace based on the reachability analysis from Section 4. Therefore, consider the control flow graph G corresponding to $B(P)$ and the set of reachable configurations CF which results from the above mentioned analysis. A counter-example trace is basically a lifting of a path through G to a sequence of statements from P. Hence, we first

need to compute a path through G. Therefore, consider some configurations cf_0 and cf_n. A path from cf_0 to cf_n through G consisting only of configurations from CF can be found by a wavefront-like approach. We iteratively compute the sets $step_k$ of configurations which can reach cf_n in exactly k steps. Those sets can be defined recursively with $step_0 = \{cf_n\}$ and $step_{k+1} = pre_{T \cup C}(step_k) \cap CF$. This sequence of sets is extended until eventually some set $step_n$ contains cf_0. Then, a path from cf_0 to cf_n through G is given by $\pi = \pi_0 \dots \pi_n$ with

$$\pi_0 = cf_0 \qquad \text{and} \qquad \pi_{i+1} = pickone(\, post_{T \cup C}(\pi_i) \cap step_{n-i-1}\,)$$

where *pickone* chooses some arbitrary complete configuration from a given set[5].

Given such a path π we can lift it to a sequence of statements from $B(P)$ by simply backtracking the translation process from Section 4. From the resulting sequence, we can generate the desired counter-example trace by, again, backtracking the abstraction process from Section 3.

Since the computed path may contain *summary edges* the resulting trace can contain function calls. As a last step, we *flatten* the trace from above and replace every function call `f();` with `f(); `τ_{sub} `return;` where τ_{sub} is a recursively computed flat trace for `f`. Finally note, when recursively descending one must not use a summary edge twice as this indicates non-termination (cf. Section 4). Listing 3 continues the example from above and gives an example trace.

```
// trace for B(ExP)
x, y = 5, 13;
swap();
// failing assert
assume(!(x > y));
assert(false);
```

```
// sub-trace for swap()
swap();
  x = x + y;
  y = x - y;
  x = x - y;
return;
```

List. 3: Spurious counter-example trace.

Validating Traces Given a flattened counter-example trace τ we want to check whether the original program P is actually able to execute statements in that particular order. This is the case if $\{true\}\tau\{false\}$ is no valid Hoare triple [Le05]. The validity of this Hoare triple can be checked by computing either the strongest postcondition of τ relative to *true* or the weakest precondition of τ relative to *false*. We choose to use the weakest precondition as it does not introduce quantifiers and tends to produce a smaller formula [He04, Le05]. That is, it remains to compute the weakest precondition $wp(\tau, false)$ according to the rules from Figure 4 and check whether it is equal to *true*.

$$wp(\tau_1;\tau_2,\varphi) = wp(\tau_1, wp(\tau_2,\varphi)) \qquad wp(assert(c),\varphi) = \varphi \wedge c$$
$$wp(x = e,\varphi) = \varphi[x \mapsto e] \qquad wp(assume(c),\varphi) = \varphi \vee \neg c$$
$$wp(x_1 \dots x_n = e_1 \dots e_n,\varphi) = \varphi[x_1 \mapsto e_1, \dots, x_n \mapsto e_n]$$
$$wp(f(),\varphi) = \varphi \text{ where every local variable is prefixed with some symbol}$$
$$wp(return,\varphi) = \varphi \text{ where one prefix is removed from every local variable}$$

Fig. 4: Rules of the weakest precondition calculus, adapted from [Le05].

[5] When using a symbolic encoding, e.g. BDDs [Sc04], multiple configurations might be "merged". Thus, π could represent multiple paths with identical length if one would skip *pickone* in the definition of π_{i+1}.

6 Abstraction Refinement

The abstraction needs refinement when a spurious counter-example trace τ from Section 5 was found. To proceed checking the original program, the abstraction needs refinement such that τ will not be produced as counter-example trace again.

An initial idea for refinement would add all formulas that where computed as weakest preconditions during the spuriosity check from Section 5. This, however, cannot be handled by our abstraction procedure as the weakest preconditions introduce copies of local variables. Hence, we need to generate new predicates that are well-scoped and well-typed. To that end, we apply the technique proposed by Henzinger et al. [He04] which is based on Craig interpolation. A Craig interpolant [Cr57] for a pair (φ^-, φ^+) is a formula ψ with (a) $\varphi^- \Rightarrow \psi$, (b) $\varphi^+ \wedge \psi$ is unsatisfiable, and (c) ψ does only contain variables common to φ^- and φ^+ .

The refinement is conducted in three steps. First, a constraint trace[6] ϑ is generated. The constraint trace ϑ is generated from τ by giving every intermediate run-time value a name. That is, a variable x is replaced with a symbolic constant $\langle x,k\rangle$ which denotes the k-th value of x. The formal rules for generating a constraint trace are given in Figure 5. The procedure requires a function *last* which maps variables to integers indicating the last write to a variable. This function is updated during the procedure to keep track of assignments and the most recent values of variables. Additionally, we utilize a function upd_f which replaces all variables x with a symbolic constant $\langle x, f(x)\rangle$.

Statement s	Constraint Trace relative to (s,f)
$S_1;S_2$	$(s';s'',f'')$ with $(s',f') =$ constraint trace relative to (S_1,f) $(s'',f'') =$ constraint trace relative to (S_2,f')
$x = e;$	$(\langle x,k\rangle = upd_f(e),\ f[x \mapsto k])$ with $k = f(x)$
$x_1 \ldots x_n = e_1 \ldots e_n;$	$(\langle x_1,k_1\rangle \ldots \langle x_n,k_n\rangle = upd_f(e_1) \ldots upd_f(e_n),\ f')$ with $k_i = f(x_i)$ $f' = f[x_1 \mapsto k_1, \ldots, x_n \mapsto k_n]$
$assert(c);$	$(assert(upd_f(c)), f)$
$f();$ $\quad \tau;$ *return*;	$(\vartheta,\ f'')$ with $(\vartheta, f') =$ constraint trace relative to (τ, f) $f'' = \begin{cases} f(x), & \text{if } x \text{ is a local variable} \\ f'(x), & \text{otherwise.} \end{cases}$

Fig. 5: Rules for generating a constraint trace, adapted from [He04].

[6] Such constraint traces where already introduced in [BR02], but the proposed refinement suffers from a similar problem as the naive approach based on weakest preconditions. So called *symbolic constants* are contained in the newly computed predicates which cannot be handled by our abstraction procedure.

Secondly, an interpolant for every intermediate position in ϑ is computed. That is, we compute the interpolants $\psi_1, \ldots, \psi_n$ where n is the number of statements in ϑ and ψ_i is the interpolant for the pair $(\varphi_i^-, \varphi_i^+)$ with

$$\varphi_i^- = \bigwedge_{j=0}^{i-1} \vartheta[j] \qquad \varphi_i^+ = \bigwedge_{j=i}^{n} \vartheta[j]$$

where $\vartheta[k]$ is the k-th statement in ϑ. Lastly, we post-process those interpolants by replacing every symbolic constant $\langle x, k \rangle$ with its corresponding variable x and extend the set of predicates with every atomic predicate contained in the post-processed interpolants. The newly added predicates are obviously well-typed and, as they originate from interpolants, are also well-scoped [He04]. The latter is due to the fact that an interpolant contains only symbols common to φ^- and φ^+, i.e. symbols that are "in scope" in φ^- and φ^+.

7 Conclusion and Future Work

In this paper we presented an integration of state-of-the-art techniques for model checking sequential recursive integer programs. To handle the infinite state space caused by integer variables and recursion, we implemented the CEGAR loop. We instantiated this loop with predicate abstraction to construct boolean programs which feature a finite data domain. To check those boolean programs, which still allowed recursion, we showed how to conduct a reachability analysis with procedure summaries. Here, we skipped some details of our actual implementation which uses binary decision diagrams (BDDs) [Sc04], an efficient data structure for handling boolean functions, to encode the generated control flow graphs. This representation allowed us to implement a variation of the procedure summaries. Instead of handling a single configuration at a time, our tool is able to compute a precise relation for multiple configurations at once. With the reachability analysis at hand, we were able to show the correctness of a program or to extract a counter-example. This counter-example was then checked for validity with weakest preconditions. A valid counter-example immediately proved the program incorrect, while a spurious one issued refinement. The refinement was conducted on the basis of Craig interpolation and enriched the abstraction with new predicates.

As our model checker was developed during a masters project, we did simplify and skip some parts. First of all, we did no exhaustive benchmarking and performance analysis. This is considered to be future work. During such an activity bottlenecks of our implementation might be identified which one could try to cure. Currently, we believe that our checker is mostly busy computing the abstraction whereas the remaining parts seem to take only a minor part of the computation time. One promising approach addressing this potential bottleneck is the so called *parsimonious* abstraction from [He04].

A further aspect, which is considered future work, too, is an improved language featuring more powerful functions. To that end, one could introduce formal parameters and (multiple) return values. There are already results for abstraction and refinement procedures for those kind of functions available in [Ba01, He04]. We believe that more powerful functions could reduce the size of programs and thus the size of the generated control flow graphs. This reduction could then lead to performance improvements.

Another area of further improvement is the counter-example generation. Currently, counter-example paths are computed by a simple breadth-first search in the k-step reachability sets. Here, state-of-the-art approaches, like proposed in [Cl95], might improve our checker.

References

[Ba01] Ball, Thomas; Majumdar, Rupak; Millstein, Todd D.; Rajamani, Sriram K.: Automatic Predicate Abstraction of C Programs. In (Burke, Michael; Soffa, Mary Lou, eds): Proceedings of the 2001 ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI), Snowbird, Utah, USA, June 20-22, 2001. ACM, pp. 203–213, 2001.

[BR00] Ball, Thomas; Rajamani, Sriram K.: Bebop: A Symbolic Model Checker for Boolean Programs. In (Havelund, Klaus; Penix, John; Visser, Willem, eds): SPIN Model Checking and Software Verification, 7th International SPIN Workshop, Stanford, CA, USA, August 30 - September 1, 2000, Proceedings. volume 1885 of Lecture Notes in Computer Science. Springer, pp. 113–130, 2000.

[BR02] Ball, Thomas; Rajamani, Sriram K.: Generating Abstract Explanations of Spurious Counterexamples in C Programs. Technical Report MSR-TR-2002-09, Microsoft Research, January 2002.

[Cl95] Clarke, Edmund M.; Grumberg, Orna; McMillan, Kenneth L.; Zhao, Xudong: Efficient Generation of Counterexamples and Witnesses in Symbolic Model Checking. In: DAC. pp. 427–432, 1995.

[Cl00] Clarke, Edmund M.; Grumberg, Orna; Jha, Somesh; Lu, Yuan; Veith, Helmut: Counterexample-Guided Abstraction Refinement. In (Emerson, E. Allen; Sistla, A. Prasad, eds): Computer Aided Verification, 12th International Conference, CAV 2000, Chicago, IL, USA, July 15-19, 2000, Proceedings. volume 1855 of Lecture Notes in Computer Science. Springer, pp. 154–169, 2000.

[Cr57] Craig, William: Linear Reasoning. A New Form of the Herbrand-Gentzen Theorem. J. Symb. Log., 22(3):250–268, 1957.

[GS97] Graf, Susanne; Saïdi, Hassen: Construction of Abstract State Graphs with PVS. In (Grumberg, Orna, ed.): Computer Aided Verification, 9th International Conference, CAV '97, Haifa, Israel, June 22-25, 1997, Proceedings. volume 1254 of Lecture Notes in Computer Science. Springer, pp. 72–83, 1997.

[He04] Henzinger, Thomas A.; Jhala, Ranjit; Majumdar, Rupak; McMillan, Kenneth L.: Abstractions from proofs. In (Jones, Neil D.; Leroy, Xavier, eds): Proceedings of the 31st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2004, Venice, Italy, January 14-16, 2004. ACM, pp. 232–244, 2004.

[JM09] Jhala, Ranjit; Majumdar, Rupak: Software model checking. ACM Comput. Surv., 41(4), 2009.

[Le05] Leino, K. Rustan M.: Efficient weakest preconditions. Inf. Process. Lett., 93(6):281–288, 2005.

[Mi67] Minsky, Marvin L.: Computation: Finite and Infinite Machines. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1967.

[Sc04] Schneider, Klaus: Verification of Reactive Systems: Formal Methods and Algorithms. Texts in Theoretical Computer Science. An EATCS Series. Springer, 2004.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

A recommended framework for anomaly intrusion detection system (IDS)

Tho Le[1]

Abstract: Signature-based solutions for Intrusion Detection are dominant in practice despite of its incapability to detect zero-day attacks. Moreover, anomaly-based Intrusion Detection Systems (IDS), a promising approach against both known and unknown attacks, are not mature for a broad productive use. Therefore, the further development of anomaly based IDS is an imperative task to strengthen security in todays networked infrastructure. This motivates a detailed study to give a structured view of problems and challenges and of the current state in this field. For this purpose, a sound analysis of current limitations and a very comprehensive survey of research papers have been conducted. In this article, a short summary of the results of the survey is given. Furthermore, the survey led to important insights into future research efforts and a proposal for a promising future IDS architecture, which is presented in this work.

Keywords: Anomaly-based IDS, survey, problems and challenges, architecture.

1 Introduction

Intrusion Detection Systems (IDS) are an urgent requirement for network security. Most of today's IDS are based on signature detection (also known as *misuse detection*). They are easy to use and have a low false positive rate. However, there is a concerning disadvantage: they are not able to detect unknown attacks like zero-day exploits. Zero-day exploits are attacks that are not publicly known and because of that, no security patches are released and no security vendor has developed signatures to detect these security issues [SY14]. Therefore, zero-day exploits pose serious threats to organizations. In [SY14], the security company *Symantec* reports that in the year 2013, 23 new zero-day vulnerabilities (out of 6787 new vulnerabilities in total) are detected. It is an increase of more than 60% in comparison to 2012 (14 zero-days), which is the highest number of zero-days since 2006.

Besides signature-based IDS, anomaly-based systems are also developed. The basic idea is that attacks are detectable by abnormal behaviors. If the anomaly-based system observed the relevant system properties, it is able to detect unknown zero-day attacks. However, there are also disadvantages in these systems: anomaly-based systems have to learn the differences between normal and abnormal behaviors. Moreover, they have a high false positive rate. These drawbacks could be explanations of why anomaly-based detection systems are not widely applied.

[1] Heilbronn, Informatik, Max-Planck-Str. 39, 74081 Heilbronn, thole020287@gmail.com

In this paper, an overview about problems and challenges as well as trends of anomaly-based detection systems is given by analyzing more than 160 studies [Le14] published in the last 3 years. Based on these trends as well as the discussed problems and challenges of anomaly-based detection, the architecture for a novel anomaly detection system is proposed [Le14].

The remainder of this paper is organized as follows: Section 2 gives an overview of today's problems and challenges of anomaly detection systems. Section 3 presents statistic information about the state of the art systems. Finally, a novel architecture for anomaly detection systems is suggested in section 4 before the conclusion in section 5.

2 Problems and challenges

Beginning in 1987, anomaly-based detection has been capturing a number of research efforts in order to develop a more sophisticated detector in protecting computing systems from compromises, especially zero-day attacks [BBK14] [BBK11] [JPP11] [GT06] [Ga08] [TSG10]. However, after more than 27 years of research efforts, such kind of systems is still missing in practical deployment. Therefore, understanding the current drawbacks is a critical and imperative task for the development in this area. Although many authors have identified ahead problems [BBK14] [BBK11] [GT06] [Ga08] [PP07] [RRR08] [TSG10] [VTN13] [SP10] [Ko11], they individually gave a piece of the puzzle, hence the work is to collect, analyze and give a more advanced and broader picture of current shortcomings. From that, the direction for further development is shown. There are four identified domains of problems existing in anomaly-based detection technique, namely hypothesis, implementation, evaluation and operation.

2.1 Hypothesis

- **Definition of normality**: is one of the key factors influencing the performance of an anomaly-based detection system. This is proven to be a complicated task since normal and malicious activities are sometimes close to each other. Furthermore, expected behaviors can be various depending on a concrete situation, i.e. local site policies of target environments [GT06] [SP10]. For example, an event is considered as normal in one environment, but may be treated as attacks in others due to their security policies.
- **Autonomous activities and intrusions:** one of the principal pitfalls is fallen in its initial premise about the interrelationship between anomalous activities and intrusions. On February 1987, Dorothy Denning originally introduced a new intrusion detection model in her paper [De87]: "security violations could be detected from abnormal patterns of system usage". This statement implicitly indicates three assumptions [GT06]: Attacks are anomalous, attacks are rare and anomalous activities are malicious. However, Gate and Taylor [GT06] questioned

and critically examined those assumptions in context of network environment to prove inappropriateness.

2.2 Implementation

- **Paradigm**: anomaly-based detection system was originally introduced to protect at single host and employed automatically to networking level, therefore, most of existing anomaly-based IDSs have not adapted properly to the modern networking paradigm [PP07] [BBK11] for both wired and wireless communication.
- **Challenges of using machine learning:** machine learning has been successfully deployed in various areas, such as voice recognition, email spam detection and it is one of the most favorable techniques in deploying anomaly-based detectors along with data mining. However, the similar success has not been found in the IDS application. Sommer and Paxson [SP10] analyzed specific characteristics in this application to find fundamental challenges in employing machine learning techniques for anomaly detection: novel attack detection and high cost of errors.

2.3 Evaluation

- **Lack of evaluation data**: lack of benchmark datasets, which can simulate realistic host and network environments, is a major issue that researchers are facing with when assessing their anomaly-based detectors [BBK14] [BBK11] [TSG10] [PP07] [SP10]. While the majority of studies were still evaluated on KDD99 dataset (published 1999) [KD14] as illustrated in section 3 and in [TSG10], it is out of date and no longer valid to represent current environments, hence such assessment results are often not reliable.
- **Lack of evaluation procedure:** many researchers [TSG10] [PP07] [SP10] believed that the inadequate evaluation process is one of the root causes that hinder anomaly-based IDS from coming into business. To such kind of critical infrastructure systems, before deploying into a real environment, it is essential to well understand and reliably evaluate its operation. Unfortunately, there is currently neither a standard framework of assessment process nor agreed evaluated metrics. Consequently, the validity of experimental results is relatively low and this may partially explain the limited success of anomaly detection in operational environments.

2.4 Operation

- **Training data:** obtaining labeled or "normal" data in real environment for anomaly training process is complicated and time consuming. Since supervised and semi-supervised algorithms require labeled and clean data respectively to learn and

build behavior profiles of targeted environments, it is critical to prepare qualified training data for optimal performance. Unfortunately, there is currently no proven process for automatically extracting those required features in a real environment. Gate and Taylor [GT06] even said, "The assumption that there exists attack-free data for training a detector outside of simulated data is not a realistic assumption".

- **High false alarm rate:** is the agreed problem shared by most of researchers for the limited widespread deployment of anomaly-based intrusion detection in real environments. Although some systems have recently achieved relatively low false alarm rates such as [MRR13] with 0.1% false alarm rate and 99.6% of detection rate, the small proportion of a large number is still a big one as mentioned in [GT06]. Therefore, reducing false positive rate is properly the most critical and imperative task toward the wide deployment of anomaly-based IDS in business.
- **Adaptability:** is the core of an effective anomaly-based detection system [GT06] [IX13]. Since intruders constantly evolve malicious behaviors to evade IDS systems, anomaly-based detectors must be able to frequently re-train their behavior profiles and adapt to new threats as well as new situations without performance influence.
- **Real time operation:** is another challenge to anomaly-based IDS [BBK14]. With the rapid increase of computer networks and heavy applications running on top of it, an anomaly-based detector is expected to process a large amount of data in a timely manner.

3 The current state of art

A comprehensive survey has been conducted to provide an up-to-date view of the current pace in anomaly-based IDS field. The survey collects 169 studies published in the last three years (2012 to 2014), from five popular digital libraries, namely: Institute of Electrical and Electronics Engineers (IEEE), Springer, Association for Computing Machinery (ACM), Google Scholar and Science Direct. From that, some interesting statistical summaries are presented with a notice that clustering, neural networks and support vector machine methods are the most employed techniques in this field. Some remarkable works can be found at [Le12], [QMG12], [MRR13], [Ab14]. More details of the survey can be referenced in the master thesis [Le14].

The survey observes a dramatic decline in the number of researches as it can be seen clearly from the chart 3.1. In 2012, 86 studies related to anomaly-based IDS were carried out, making this number the highest in the last 3 years. In 2013, the number has declined to 51 and continued to decrease to 26 reports published in 2014. This linear drop over the last three years may be suggested by many reasons, one of which could imply for the reduction of interest in this area, especially in the situation that this technique is still considered as immature after 27 years of continuous development.

In general, an anomaly-based detector can be trained to build a normal behavior profile via three ways, namely supervised, semi-supervised, unsupervised. Both supervised and semi-supervised systems require labeled and clean training data to build normal profiles respectively. However, unsupervised approaches can be trained with raw data. Those unsupervised systems require minimum effort from security experts since they are able to learn acceptable behaviors and adapt themselves to behavior drifts from targeted environments. Therefore, it is anticipated to be the most widely adopted and this is supported by the survey's result with 37% of total research works as illustrated in the figure 3.2. With less than 2%, supervised anomaly detectors also capture a considerable amount of interest, as they tend to produce less false alarms while gaining higher detection accuracy than unsupervised ones. Moreover, the semi-supervised mode is responsible for the smallest contribution in the anomaly field. This can be explained by the facts that: 1) exposing such IDS systems to only normal traffic is also a challenge since it requires intensive scanning and screening to remove various kind of hidden attacks, 2) semi-supervised systems are less efficient than supervised ones in terms of detection accuracy and false alarm generation.

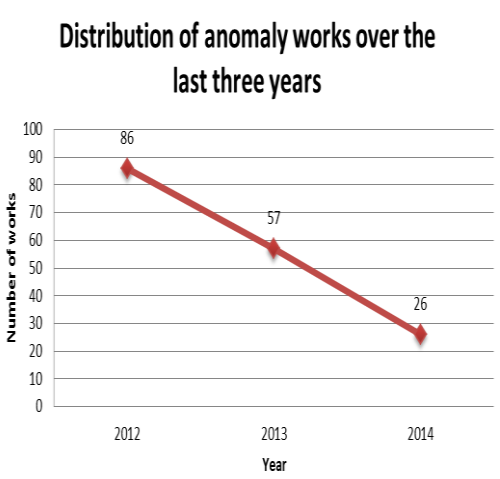


Fig. 3.1: Distribution of anomaly works [Le14]

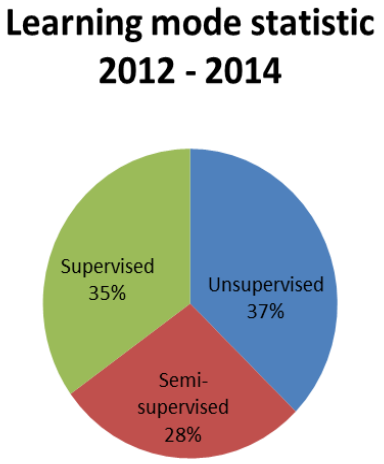


Fig. 3.2: Learning mode statistic [Le14]

Assessment is an integral part in developing an IDS tool to evaluate its reliability, efficiency and effectiveness before applying in real life scenarios. This phrase cannot be done without qualified evaluation datasets. Therefore, it would be necessary to survey datasets on which scientists are working to develop their detectors. As illustrated clearly in the figure 3.3, more than half of researches were experimented with DARPA families, including DARPA [LL14], KDD99 [KD14] and NSL-KDD [NS14], in which KDD99 is the most famous dataset, constituting 35% of total experiments. The other datasets are contributing much smaller proportions. While about a quarter of anomaly evaluations were performed with synthetic and real datasets individually, yet all other benchmarked

datasets, such as Sendmail [CS14], Kyoto 2006+ [TA14] etc. are only accounting for 20% of total experiments conducted in the survey.

The pie chart 3.4 illustrates the proportion of under-lying algorithms contributing to the development of anomaly-based IDS during the three-year period. As expected, machine learning based and data mining based techniques are the main stream of favor, being responsible for more than 50% of total implementations. Slightly below those two, with 24%, the survey observes a new trend of combining multiple techniques within or across categories to improve detection performance called "combination". This new approach comes from the fact that each technique contains its own advantages and disadvantages. Therefore, a proper combination of those techniques could result in a better detection. Furthermore, statistical based algorithms are the least favorite technique used in during the last three years (21%). Noticeably there is no work applying knowledge-based techniques.

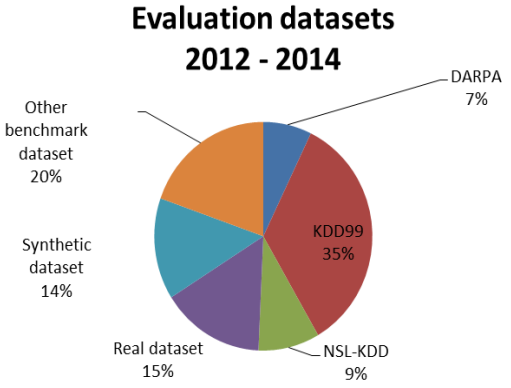


Fig. 3.3: Evaluation datasets [Le14]

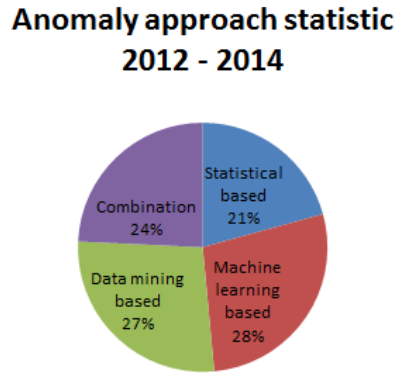


Fig. 3.4: Anomaly approach statistic [Le14]

4 The Recommended architecture

In order to provide meaningful recommendations, it is necessary to analyze carefully requirements on an anomaly IDS system as below:

- **Efficient detection performance**: detecting and alarming any intrusions are the primary functions of an IDS system, therefore, the first and most important requirement is to detect attacks as many as possible with a minimum number of false alarms. Furthermore, delay time for each decision must be small enough in the context of optimal computing resource usage.
- **Classification**: the conventional purpose of anomaly detection is to distinguish between normal and anomalous activities. However, in many cases, alerting anomalous events are not helpful enough to administrators, since they have to spend a considerable amount of effort and time on investigating types of attacks

before deciding appropriate responses. Therefore, it is more useful to alarm anomalous events along with their belonging categories, such as Probe, D.o.S etc.

- **Easy deployment**: this kind of systems should be straightforward to install and operate in targeted environments. Supervised and semi-supervised learning approaches require labeled and normal data for training process, which makes it complex to implement in business environments. Ideally, training phase in supervised and semi-supervised systems should be either bypassed or replaced by another process placing no stress on training data like in unsupervised approach.
- **Adaptability**: anomaly-based IDS must be able to adapt itself to changes of dynamic environments such as enterprise networks.
- **Real time operation**: It must be able to process a large volume of data in a timely manner.

Since signature and anomaly detections are two opposite directions that advantages of this are weaknesses of other, the idea of combining those two into a hybrid model seems to be very promising and that is also the cornerstone in the recommended architecture. Basically, there are three ways to combine signature and anomaly approaches together. Anomaly and signature models can run in parallel and two outcomes are merged or selected for a final decision, otherwise either anomaly or signature can be implemented on top of other. In this recommendation, the latter approach is chosen. The signature module is implemented in the first layer since it provides a better integration between the signature and anomaly modules. In this context, the signature module acts as a filter to remove all known attacks and leaving only normal and unknown attacks to the subsequent layer. Furthermore, the anomaly module inherits detection results from the upper module for the training process. Particularly, those detected known attacks along with the classification results from the unsupervised anomaly layer are used as labeled training data for a supervised algorithm. The general recommended architecture is clearly illustrated in the figure 4.1:

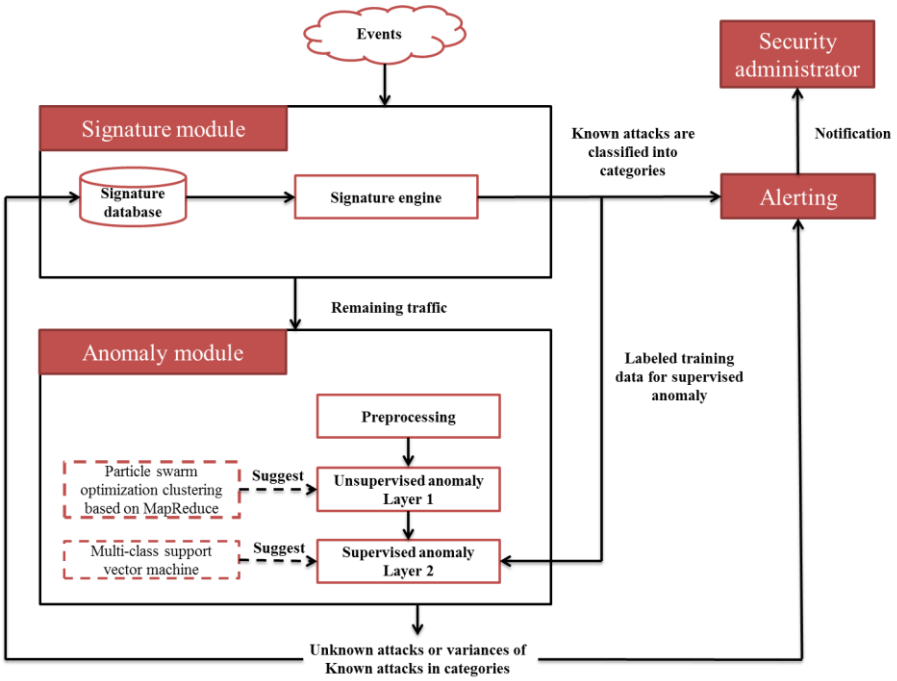


Fig. 4.1: The recommended architecture

- Events, which can be network traffic data or system call logs, are input into the signature module for the detection process.
- The signature database is utilized to detect known attacks in input data and classify them into predefined categories before sending to alerting module for notification. At the same time, those known attacks are removed from original data, leaving only normal and properly zero-day attack data to the anomaly module.
- The remaining traffic will be firstly preprocessed to select necessary features and normalize them for later detection process. The reduced data are then applied into the first layer, which employs an unsupervised algorithm to group instances into two categories: normal and anomalous. In order to process a large volume of data in a real time manner, it is proposed to employ parallel particle swarm optimization clustering algorithm based on MapReduce methodology [AL12] as inspired by the work [AL13]. Then results of each cluster in the first layer are forwarded to the supervised layer for final classification decisions. In this context, the multiple-class SVM technique [Le12] is suggested since it achieved the best performance in the conducted survey. The final detection of the anomaly module can be unknown attacks or variances of known attacks, which will be then sent to the alerting module for notification. At the same time, these will be updated to

signature database so that in next loops, those malicious activities are filtered in the signature layer.

Since the signature detection is not the focus of the thesis, its detailed operation is not covered in this section. Instead, the anomaly module is further elaborated with two working modes: training and operation.

- In the training mode as shown in figure 4.2, raw data, which can be normal or unknown attacks from the output of the upper signature module, are input into the unsupervised algorithm to build behavior profiles in forms of normal and anomalous clusters. For each cluster, a corresponding instance in the supervised layer is created as a second level of classification for sound detection decisions. Since the supervised algorithm requires labeled data for training, clustering results and known attacks detected in signature module are used to build classification models.

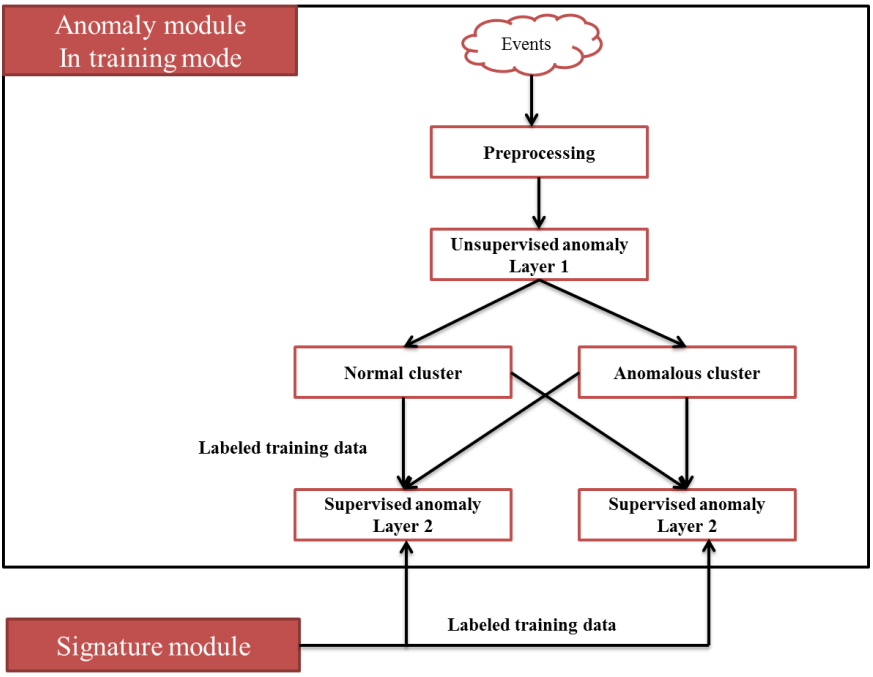


Fig. 4.2: Anomaly module in training mode

- In the operation mode, classification results of each cluster in layer 1 are forwarded to each corresponding instance in the supervised layer for final classification decisions as presented in figure 4.3.

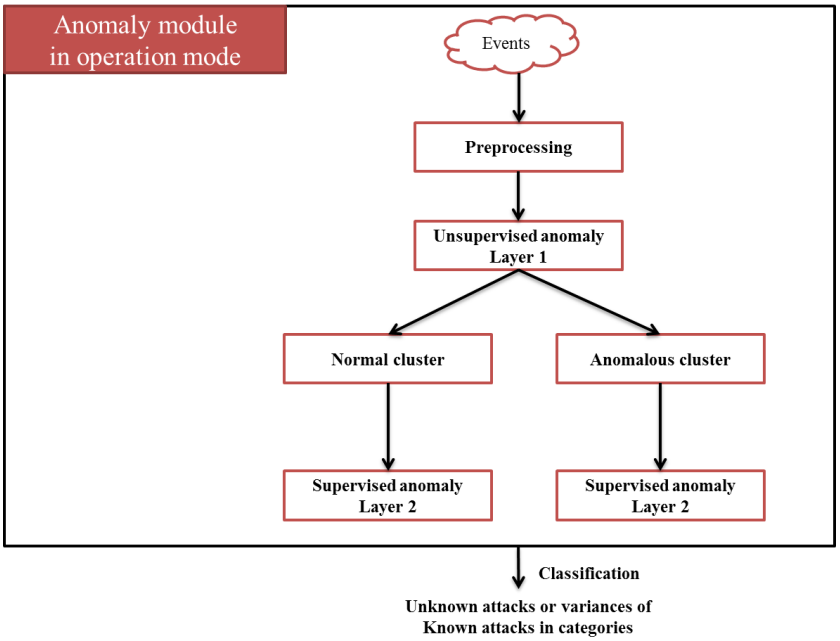


Fig. 4.3: Anomaly module in operation mode

5 Conclusion

In this paper, three main contributions have been presented:

- An advanced view on the current problems and challenges of the anomaly-based IDS.
- An overview about current trends and technologies that is based on the survey of more than 160 published studies.
- A recommended architecture is proposed for a more efficient IDS.

However, there are some limitations of the work: the suggested architecture is actually only proposed in literature without having a prototype to check its performance and quality of detection. Furthermore, because of the variety of publications, the survey reviewed only a short period of time (2012-2014); therefore, it may not reflect the entire current trends and tendencies. It is also important to note that the analyzed studies present only the state of art and not the state of practice.

For future works, I consider further researches to improve the understanding of the limitations of current technologies employed as very important. Existing problems, which limit the potential of the technique, should be analyzed in more detail. In addition,

an implementation of the suggested architecture is undertaken to evaluate its efficiency in practice. On the other hand, the current state of art of anomaly-based IDS should be reviewed in a longer period for a better overview of research efforts and technology trends.

Acknowledgements

I would like to thank my supervisors, Prof. Dr. Friedbert Kaspar and Mr Fabian Berner, without their instructions and dedicated support through each step of the process, this study would not ever be accomplished.

References

[Ab14] Abuadlla.Y. et.al. : Flow-based Anomaly Intrusion Detection System using Two Neural Network Stages: Computer Science and Information Systems 2014, pp. 601 – 622, 2014.

[AL12] Aljarah, I; Ludwig, S: Parallel Particle Swarm Optimization Clustering Algorithm based on MapReduce Methodology: Nature and Biologically Inspired Computing, Mexico City, pp. 104 - 11, 2012.

[AL13] Aljarah, I; Ludwig, S: MapReduce Intrusion Detection System based on A Particle Swarm Optimization Clustering Algorithm: Evolutionary Computation, Cancun, pp. 955 - 962, 2013.

[BBK11] Bhuyan, M; Bhattacharyya, D; Kalita, J: Survey on Incremental Approaches for Network Anomaly Detection. International Journal of Communication Networks and Information Security 11/11, pp. 226 - 239, 2011.

[BBK14] Bhuyan, M; Bhattacharyya, D; Kalita, J: Network Anomaly Detection: Methods, Systems and Tools. Communications Surveys & Tutorials 02/14, pp. 303 - 336, 2014.

[CS14] UNM Synthethic Sendmail Data, http://www.cs.unm.edu/~immsec/data/synth-sm.html, as of 10.12.2014.

[De87] Denning, D: An Intrusion Detection Model: IEEE Transactions on Software Engineering, pp. 222 - 232, 1987.

[Ga08] García-Teodoro, P. et.al. : Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges. Computers & Security 08/08, pp. 18 - 28, 2008.

[GT06] Gates, C; Taylor, C: Challenging the Anomaly Detection Paradigm: A Provocative Discussion: 06 Proceedings of the 2006 Workshop on New Security Paradigms, New York, pp. 21 - 29, 2006.

[IX13] Ippoliti, D; Xiaobo, Z: A Self-Tuning, Self-Optimizing Approach for Automated Network Anomaly Detection Systems: Proceedings of The 9th International Conference on Autonomic Computing, pp. 85 - 90, 2013.

[JPP11] Jyothsna, V; Prasad, V; Prasad, K: A Review of Anomaly based Intrusion Detection Systems. International Journal of Computer Applications 08/11, pp. 26 - 35, 2011.

[KD14] KDD Cup 1999 Data, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, as of 10.12.2014.

[Ko11] Koch, R: Towards Next-Generation Intrusion Detection: Cyber Conflict (ICCC), 2011 3rd International Conference,Tallinnn, pp.1 - 18, 2011.

[Le12] Le, V. et.al. : Network Intrusion Detection based on Multi-class Support Vector Machine: ICCCI'12 Proceedings of the 4th International Conference on Computational Collective Intelligence: Technologies and Applications, pp. 536-543, 2012.

[Le14] Le.T: Anomalie-based Security Analysis, Master of Science Thesis, 2014.

[LL14] DARPA Intrusion Detection Evaluation, http://www.ll.mit.edu/ideval/, as of 10.12.2014.

[MRR13] Muniyandi, A; Rajeswari, R; Rajaram, R: Network Anomaly Detection by Cascading K-means Clustering and C4.5 Decision Tree Algorithm: International Conference on Communication Technology and System Design 2011, pp. 174 - 182, 2013.

[NS14] The NSL-KDD Dataset, http://nsl.cs.unb.ca/NSL-KDD/, as of 10.12.2014.

[PP07] Patcha, A; Park, J: An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends. Computer Networks 08/07, pp. 3448 - 3470, 2007.

[QMG12] Qazanfari, K; Mirpouryan, M.S; Gharaee, H: A Novel Hybrid Anomaly Based Intrusion Detection Method: Telecommunications (IST), Tehran, pp. 942 – 947, 2012.

[RRR08] Ringberg, H; Roughan, M; Rexford, J: The Need for Simulation in Evaluating Anomaly Detectors: ACM SIGCOMM Computer Communication, New York, pp. 55 - 59, 2008.

[SP10] Sommer, R; Paxson, V: Outside the Closed World: On Using Machine Learning for Network Intrusion Detection: IEEE Symposium on Security and Privacy, Oakland, pp. 305 - 316, 2010.

[SY14] Internet Security Threat Report 2014, http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v19_21291018.en-us.pdf, as of 10.12.2014.

[TA14] Traffic Data from Kyoto University's Honeypots, http://www.takakura.com/Kyoto_data/, as of 10.12.2014.

[TSG10] Tavallaee, M; Stakhanova, N; Ghorbani, A: Toward Credible Evaluation of Anomaly-based Intrusion-Detection Methods. Systems, Man, and Cybernetics 09/10, pp. 516 - 524, 2010.

[VTN13] Viswanathan, A; Tan, K; Neuman, C: Deconstructing the Assessment of Anomaly-based Intrusion Detectors, Springer Berlin Heidelberg, 2013.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Java type inference as an Eclipse plugin

Andreas Stadelmeier[1]

Abstract: While in Java 5.0 generics have been introduced in Java 8 the language has been expanded by lambda expressions. The very popular feature in functional programming languages, type inference, is introduced in Java, but only in a very restricted form, similar as in other object-oriented languages. This paper presents an Eclipse plugin which allows writing Java 8 programs without any type annotation. A type inference algorithm determines all possible types which the Eclipse plugin presents to the user afterwards. So the user can select the desired type. This plugin allows to determine most general types, which are often not obvious. The programs therefore become more reuseable, which is the goal of any software development.

Keywords: Code generation, language design, program design and implementation, type inference, type system

1 Introduction

Java has been extended by two features that are well-known from functional programming languages. In Java 5.0 generics have been introduced while in Java 8 lambda expressions have been established. A further very convenient feature from functional programming languages, the type inference, is included only in a very restricted form. Let us consider the following example.

```
interface Fun1<R, T> { R apply(T arg); }
interface Fun2<R, T1, T2> {R apply(T1 arg1, T2 arg2); }

class Matrix {
  Fun1<Fun1<Matrix,Fun2<Matrix,Matrix,Matrix>>,Matrix>
    op = (m) -> (f) -> f.apply(this, m);
}
```

The function `op` applies a given function `f` (second argument) by the method `apply` to `this` and `m` (first argument).

[1] Baden-Württemberg Cooperative State University Stuttgart, Department of Computer Science, Florianstraße 15, D–72160 Horb, a.stadelmeier@hb.dhbw-stuttgart.de

If we consider the declaration of `op` we recognize that the arguments `m` and `f` are untyped. This is possible due to Java 8's restricted type inference system. But the variable `op` must be typed. As a lambda expression is assigned to `op` and as no real function-types are included in Java 8, a functional interface is the compatible target type of `op`. In our example we declare parametrized functional interfaces `Fun1` and `Fun2`, which correspond to real function-types.

If we consider the function `op` in detail we will see that the given type is not the most general. The type

```
<A,B>[2]Fun1<Fun1<B, Fun2<B, Matrix, A>>, A>
```

is also correct but it is more general. Unfortunately, there are further correct typings, where it is not obvious, which is the most general one.

Our type inference system allows to declare

```
class Matrix extends Vector<Vector<Integer>> {
    op = (m) -> (f) -> f.apply(this, m);
}
```

where no additional functional interfaces must be declared and no type for `op` must be given. The system determines all most general types[3] as functional interfaces of the form `Fun`*N*. As there are more than one most general type our Eclipse plugin allows the user to select the favored type.

This approach supports the programmer significantly, as no functional interfaces as target types of lambda expressions have to be defined, types of complex expressions with generics and wildcards need not to be detected and writing reusable code is assisted by providing most general types.

At the moment, object-oriented languages mostly use local type inference (e.g. Scala [Od14]). Local type inference determines types by a combination of local type propagation, from adjacent nodes in the syntax tree, and local constraint solving, rather than by global constraint solving [PT00, OZZ01]. In languages with local type inference only a few type annotations can be omitted.

Our approach is based on global constraint solving and allows to omit all type annotations. Nevertheless the type inference system is a practical application, as the user is supported by an intelligent Eclipse plugin.

[2] The declaration of generic fields is not allowed in Java. In our extension we allow this, if the field represents a function given by a lambda expression.

[3] A type is considered most general when there is no type with a more general meaning. But often there are multiple most general types (comparable to multiple minimal elements in a partial ordering, which are not a minimum). A most general type is therefore normally not a principal type.

The paper is organized as follows. The next section gives a brief overview of type inference following the description of our type inference algorithm. The third section presents our Eclipse plugin followed by a summary and an outlook at the very end.

2 Type inference

Since the eighties much research has been done on type inference in programming languages. The base of nearly all type inference algorithms is the algorithm $\mathscr{W}$, that was given by Damas and Milner for ML [DM82].

The Hindley–Milner type inference algorithm determines for an untyped program a principal type. Principal type means that there are other correct typings, which are less general. In the Hindley–Milner approach less general means, there is substitution, that maps the principal type to the less general type.

For object-oriented programming languages with subtyping and parametric polymorphic types (generics) *local type inference* [PT00, OZZ01] is used. Local type inference is given by two techniques: First, type parameters in a function application are inferred from the function's value parameters by solving a constraint system. Second, it propagates known types down the syntax tree in order to infer some types of formal value parameters and provide additional guidance to type parameter inference. The Java type inference is a form of local type inference, where, however, some actually unnecessary type annotations are demanded. Consequently Scala uses local type inference.

2.1 Related Work

We considered until now type inference for different Java versions. In [Pl07] we presented a type inference algorithm for Java 5.0 with generics and wildcards. In [Pl11] we gave a type inference algorithm for Java with lambda expressions and real function types, that was a first idea for Java 8 [La10]. The type inference algorithm, which the Eclipse plugin is based upon, presented in this paper, is an improved extension of [Pl11] adapted to the functional interfaces of Java 8 [Pl14].

2.2 Comparision with other type inference algorithms

Type inference is already available in Java and other similiar programming languages. This chapter shows the differences between our plugin and other type inference systems which seem analogical to our approach.

2.2.1 Comparision with local type inference in Java 8

Java 8 is able to infer the parameter types of a lambda expression. Also the type of a lambda expression itself is inferred by the compiler. This is possible due to the fact that

there is a target type of the lambda expression in the environment. If a lambda expression is assigned to a variable or passed as a parameter the compiler uses the target type to determine the type of the lambda expression. Subsequently, the parameter types as well as the return type of the lambda expression are provided by the target type.

This does not work in case a method is overloaded with two functional interfaces, which are structurally the same. In Fig. 1 the interfaces `F1` and `F2` are structurally the same. The Java compiler is not able to determine which method has to be invoked: `m(F1 p)` or `m(F2 p)`.

```
interface F1{String apply();}
interface F2{String apply();}

class Test{
    String m(F1 p){ return p.apply(); }
    void m(F2 p){ }
    String main(){ return m(()->{return "Test";}); }
}
```

Fig. 1: Compilation error with Java 8

The global type inference algorithm described in this paper is able to infer the right type for the lambda expression in Fig. 1. It considers more than the target type to determine the missing type and is able to include and use more of the given information. Due to the fact that the `main` method returns a `String` and just one of the overloaded `m` methods does as well, only the type `F1` suits for the lambda expression in this context.

2.2.2 Comparision with local type inference in Scala

In Scala the local type inference [PT00, OZZ01] is implemented. Local type inference often allows omitting types of variable declarations and returns types of methods. Furthermore the types of instantiated generic classes or polymophic methods can be inferred, too. In the Scala tutorial[4] the mechanism is explained. See some examples in the following:

```
val x = 1 + 2 * 3        //the type of x is Int
val y = x.toString()     //the type of y is String
def succ(x: Int) = x + 1 //method succ returns Int values

class MyPair[A, B](x: A, y: B);

val p = MyPair(1, "scala") // type: MyPair[Int, String]
```

In contrast recursive defined functions must be typed explicitly. The following example would cause an error:

```
def fac(n: Int) = if (n == 0) 1 else n * fac(n - 1)
```

[4] `http://docs.scala-lang.org/tutorials/tour/local-typeinference.html`

In contrast to the considered type inference approaches for Java 8 and Scala in the approach presented by this paper all types can be inferred, automatically.

2.2.3 Comparision with Xtend

The Eclipse project also worked on a way to improve the Java programming language and created their own programming language named Xtend [Xt15]. Xtend is a modified version of Java. One of its main features is the possibility to omit type declarations in the code. The Xtend compiler inserts missing types into the source code before the actual compilition through the Java compiler. Therefore Xtend uses a local type inference algorithm which behaves similiar to the type inference used in Java 8. This approach of using a type inference algorithm to produce a less noisy Java is similiar-sounding to the type inference plugin described in this paper. But a comparision between Xtend and our Eclipse plugin shows the differences and capabilities of the respective type inference algorithms.

Xtend is not capable of determining types of recursive method calls. In the following example the type of *method* can not be determined by Xtend wherefore *Object* is assigned as the return type.

```
def method()
    if(...) return 1;
    else return test2();
```

Additionally Xtend is bound to a few constraints when it comes to infer the type of a local variable. Local variables must be declared with a initial value and only the type of the initial value is considered by the type inference algorithm of Xtend. Therefore the following Xtend code snippet generates a type error.

```
var variable = new Integer(1);
    variable = new Object(); //Error: cannot convert from Object to Integer
```

Our Eclipse plugin is able to infer the usecases given above due to a global type inference algorithm. Another difference between those two Java derivatives is the amount of modifications and extensions of the Java language. Xtend brings new keywords and the ability to omit unnecessary keywords and types. Our subset of Java sticks to the Java syntax and only adds the functionality of a global type inference algorithm. It supports the programmer in finding the most general type and writing less redundant and more reusable code.

3 Type inference algorithm

This chapter descripes our type inference algorithm which is also used by the Eclipes plugin. First a outline of the procedure of the algorithm is given following the introduction of relevant data structures.

3.1 Procedure of the global type inference algorithm

The procedure of the type inference algorithm consists out of the following parts.

Gathering constrains: A deep search algorithm crawls through the abstract syntax tree. It completes the tree by adding a type variable (type placeholder) for each missing type. Additional it generates constraints for every instruction and expression.

Unify: The algorithm described in [Pl09] unifies the gathered constraints. This process either succeeds by returning a set of unified type constraints or fails returning nothing.

Instantiating a solution: The type unification has multiple solutions. Each solution represents a possible replacement of the type placeholders with actual types. Therefore one solution has to be selected and instantiated to the type placeholders in the abstract syntax tree. This can be done by the user of our Eclipse plugin.

3.2 Type Placeholder

Type placeholders are used to fill out missing types in the syntax tree. They are treated like normal types during constraints generation. The unify algorithm tries to find the most general type solution for the type placeholders within the constraints set. This can either be a concrete type or a type constraint. For example the type placeholder has to be the subtype of another type. The type inference plugin later on has to fill in the unified type solutions into the inferred source code. Therefore every type placeholder keeps a reference to its parent node in the syntax tree. So the plugin is able to determine the associated statement in the source file to a given type placeholder.

3.3 Functional interfaces

In Java 8 a lambda expression is treated as an instance of a functional interface. A functional interface is an interface with only one abstract method. To type every possible lambda expression the typeless Java language needs a new set of functional interfaces called `FunN`. For all $N \in \mathbb{N}$ there is an interface

```
interface FunN<R,T1, ... , TN> {
  R apply(T1 arg1, ... , TN argN);
}
```

For each lambda expression exists a `FunN` interface which is suitable as a type. These functional interfaces play the role of function types.

3.4 Overloading

The type inference algorithm is not always able to determine a singular set of constraints out of a method call. In a typeless Java source code the only information about a method invocation is its name and the number of its parameters. In some cases multiple methods are matching with the given information. Each match implies different constraints for the types of the expressions (e.g. the given method parameters or the return type) involved in the method call. So the algorithm regards every possible case and builds several constraint sets. The cartesian product over these constraint sets represents the set of potential typings.

This means, that besides the unification, overloading is the second cause of multiple type solutions [SP15].

4 Plugin description

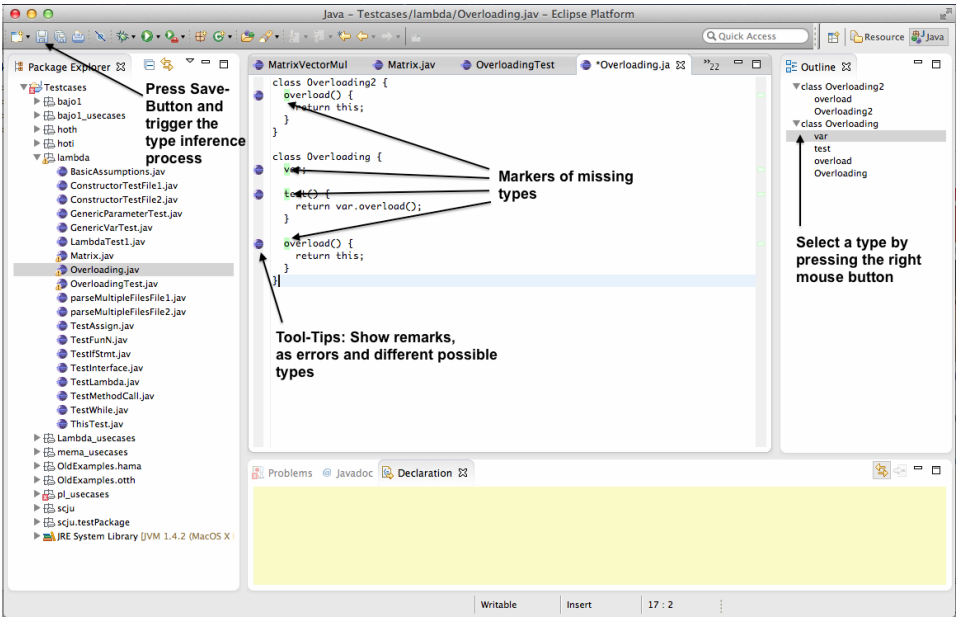


Fig. 2: Eclipse plugin

The goal of our type inference algorithm is to generate valid Java class files out of a typeless Java source code, which is achieved in three steps. First a given source file has to be inferred by the type inference algorithm. Subsequently the plugin patches the untyped statements in the source file by plugging in the inferred types. Afterwards the standard Java compiler is able to compile the resulting source code. This imposes the restriction that only valid typed Java source files are accepted by the standard Java compiler. But in most cases the type inference algorithm described in this paper generates more than one solution for the typing of a given untyped Java source code. To generate Java byte-code from the given typeless Java source code one of the infered type solutions has to be chosen.

A solution for the illustrated problem is to let the user select one of the given typings. This feature is implemented as a Eclipse plugin. In Fig. 2 the plugin is shown. The user workflow is given in Section 4.4. Section 4.1 describes the implementation, while Section 4.3 explains the user interface.

4.1 Implementation

This section describes the principles and processes of the Eclipse plugin implementation. Every time the type inference plugin is triggered on a Java document, it processes the request in the following way:

1. The source code of the document is processed by the type inference algorithm. This generates type solutions for every missing type in the syntax tree.
2. Every position in the document, where a type can be plug in gets marked by an annotation.
3. Not every type in the syntax tree is associated with a type declaration in the corresponding Java source code. Types appear in Java source code only at method declarations, field declarations and local variable declarations. Only the inferred types representing a type at the named positions are important for the Eclipse plugin. Therefore the plugin concatenates the inferred types with the previously generated markers. This is done by filtering out the types which are actually insertable into the source code and attach them to their respective markers.

4.2 Deploy type solution

The outcome of the type inference process is a list of possible type solutions.

Afterwards the plugin has to provide an interface to the user to choose one of the solutions. The implementation has to fulfill the following constraints:

Select a type for each untyped statement A source file can contain multiple statements with a missing type declaration. The user should be able to choose a type for each of those positions in the source code.

Annotations for missing types Every place a type is missing in the document an annotation is added. This annotation provides the possible types.

Integrate into user workflow Editing the document between two type selections should be possible. Every type insertion must generate a valid, although partial typeless, Java program. Therefore after each type selection the user is able to edit the source code and afterwards the other missing types can be inferred by the type inference algorithm.

Final outcome must be valid Java source code When the user selects one of the proposed types of a marker the plugin edits the document by placing in the selected type. When every marker is resolved all the missing types are patched. This process renders a valid Java source file.

There is a challenge with generic type variables. Sometimes a most general type solution contains a generic type variable. In Java a type variable must be introduced by a generic class declaration, a generic method declaration or a generic constructor declarations. Therefore a generic type variable can not be solely placed into the source code. Additionally, the declaration of the type variable has to be placed in the corresponding method or class declaration.

The source code given in Figure 3 is an example for the described practice. The types of the method `ident` and its parameter `m` are undefined. The method should be applicable to arguments of all types and the type of the parameter `m` has to extend the return type of the method `ident`. If an user triggers the plugin to inject one of the two missing types an additional declaration of a generic type variable is added. To plug in all missing types two additional type variables must be introduced (see Figure 4).

```
class GenericExample {
    ident( m ){
        return m;
    }
}
```

Fig. 3: Example: Type insertion with generics

```
class GenericExample {
    <A, B extends A> A ident( B m ){
         return m;
    }
}
```

Fig. 4: Example 3 after the plugin set in all missing types

4.3 User interface

Eclipse is a modular environment. Nearly every part of the application is replaceable or extensible through plugins. Therefore a plugin can introduce a new file type (*.jav*-file) for typeless Java code editing.

Our plugin installs only minimalistic modifications to the look and feel of Eclipse. No buttons or menu items are added. Only a new editor is used for the *jav*-files. This is needed to show different error markers and syntax highlighting than the Java editor provided by Eclipse. The editor for *jav*-files acts like the standard Eclipse Java editor.

The following controls have been added to Eclipse user interface.

Markers and annotations Our plugin displays a marker for every missing type in the *jav*-editor. These markers provide a *Quick fix* which offers the inferred types as solutions to the user.

Outline view The outline view shows an additional overview of the code displayed by the *jav*-editor.

4.4 User workflow

This section describes the workflow when using our type inference plugin for Eclipse on the basis of the example shown in Figure 5. The example source code shows a source file which defines two classes. None of the types for the methods and fields of these classes are specified. The process of inserting the missing types into the source code is running as follows (cp. Figure 2).

By saving a document in Eclipse the type inference plugin is triggered. Afterwards markers are shown in the document view which signal the possibility for a type insertion. For the field variable `var` and the method `test` in the class `Overloading` two types are possible. The annotations at this points in the source code provide a context menu with the two possible type options: `Overloading` and `Overloading2`. If the user selects one of the proposed types the plugin places the type into the source code of the document. This is done via the marker annotations displayed inside the *jav*-editor or the outline view (cp. Figure 2). After a missing type is patched the number of possible type solutions decreases. After the type of either the method `test` or the field `var` has been determined only one primary type solution remains. Figure 6 shows the outcome after selecting the type `Overloading2` for the field `var`. With only one possible type solution remaining the source code is ready for compilation. The user need not insert those missing types. Afterwards Java class files can be generated.

```
class Overloading2 {
  overload(){
    return this;
  }
}

class Overloading {
  var;

  test(){
    return var.overload();
  }
  overload(){
    return this;
  }
}
```

Fig. 5: Example for Overloading

```
class Overloading2{
  overload(){
    return this;
  }
}

class Overloading {
  Overloading2 var;

  test(){
    return var.overload();
  }
  overload(){
    return this;
  }
}
```

Fig. 6: Example for Overloading (Figure 5) after the type for `var` has been set in

5 Summary and future work

In well-known object-oriented languages as original Java or Scala, type inference is only considered in a local framework. The reason is that there was no practical implementation for a global approach. This is caused by the fact that there are often multiple solutions and it is difficult to define a principal type. This paper presents a global approach for type inference in Java 8. First it introduced a set of functional interfaces Fun*N*, that represent the function types. Then it described the algorithm, which determines all most general solutions. Finally an Eclipse plugin is presented, which allows the user to write a Java program without type definitions. The Eclipse plugin determines the solutions for all left out types in the Java program. If there is more than one solution the position is marked and the user can select the favored type. The plugin supports Java programming significantly, as no functional interfaces must be declared for the typing of lambda expressions, complex types with generics and wildcards need not to be detected and as the inferred types are most general the code is best possible reuseable.

5.1 Future work

Currently our type inference algorithm mostly produces multiple most general types for the same field or method when used on common Java source code. As long as no single principal type is inferred the user has to select one of the proposed type solutions. In [Pl08] we described a resolving strategy which results in an intersection type, that is principal. In future we will improve our system, such that the inferred type is the prinicipal intersection type. Additionally, the system will generate byte-code for the fields and methods typed by principal intersection types, which means for the user that the type selection would no longer be necessary.

References

[DM82] Luis Damas and Robin Milner. Principal type-schemes for functional programs. *Proc. 9th Symposium on Principles of Programming Languages*, 1982.

[La10] Lambda. Project Lambda: Java Language Specification draft, 2010. Version 0.1.5.

[Od14] Martin Odersky. *The Scala Language Specification Version 2.9*, May 2014.

[OZZ01] Martin Odersky, Christoph Zenger, and Matthias Zenger. Colored local type inference. *POPL 2001 Proc. 28th ACM Symposium on Principles of Programming Languages*, 36(3):41–53, 2001.

[Pl07] Martin Plümicke. Typeless Programming in Java 5.0 with Wildcards. In Vasco Amaral, Luís Veiga, Luís Marcelino, and H. Conrad Cunningham, editors, *5th International Conference on Principles and Practices of Programming in Java*, volume 272 of *ACM International Conference Proceeding Series*, pages 73–82, September 2007.

[Pl08] Martin Plümicke. Intersection Types in Java. In Luís Veiga, Vasco Amaral, Nigel Horspool, and Giacomo Cabri, editors, *6th International Conference on Principles and Practices of Programming in Java*, volume 347 of *ACM International Conference Proceeding Series*, pages 181–188, September 2008.

[Pl09] Martin Plümicke. Java type unification with wildcards. In Dietmar Seipel, Michael Hanus, and Armin Wolf, editors, *17th International Conference, INAP 2007, and 21st Workshop on Logic Programming, WLP 2007, Würzburg, Germany, October 4-6, 2007, Revised Selected Papers*, volume 5437 of *Lecture Notes in Artificial Intelligence*, pages 223–240. Springer-Verlag Heidelberg, 2009.

[Pl11] Martin Plümicke. Well-typings for Java$_\lambda$. In *Proceedings of the 9th International Conference on Principles and Practice of Programming in Java*, PPPJ '11, pages 91–100, New York, NY, USA, 2011. ACM.

[Pl14] Martin Plümicke. More type inference in Java 8. In *Perspectives of Systems Informatics - 9th International Andrei Ershov Memorial Conference, PSI 2014, St. Petersburg, Russia*, pages 168–174, 2014. Preliminary Proceedings.

[PT00] Benjamin C. Pierce and David N. Turner. Local type inference. *ACM Trans. Program. Lang. Syst.*, 22(1):1–44, 2000.

[SP15] Andreas Stadelmeier and Martin Plümicke. Adding overloading to Java type inference. In *Tagungsband der Arbeitstagung Programmiersprachen (ATPS 2015)*, volume Vol-1337, pages 127–132. CEUR Workshop Proceedings (CEUR-WS.org), 2015.

[Xt15] Xtend-language, april 2015. http://www.eclipse.org/xtend/.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Evaluation der Leistungsfähigkeit ausgewählter Mutationstestwerkzeuge

Lea Kristin Gerling[1]

Abstract: Mutationstests dienen zur Qualitätsbestimmung von Softwaretests. Um die Ausführung dieser Tests zu automatisieren, gibt es eine Reihe von Mutationstestwerkzeugen. Diese Arbeit beschäftigt sich mit den öffentlich verfügbaren Werkzeugen für die Programmiersprache Java und dem Testframework JUnit. Sie vergleicht die Werkzeuge MuJava, Jumble, Pitest, Judy und Major hinsichtlich ihres Funktionsumfangs, ihrer Leistungsfähigkeit und ihrer Praxistauglichkeit miteinander. Die Ergebnisse dieser Evaluation dienen als Entscheidungshilfe bei der Auswahl eines Werkzeugs zur Testverbesserung.

Keywords: Mutationstest, Mutationsanalyse, Mutationstestwerkzeug, Evaluation.

1 Einleitung

Softwaretests sind ein wichtiger Bestandteil der Softwareentwicklung. So spart frühzeitiges Testen nicht nur Geld, da ein Fehler umso mehr Kosten verursacht, je später er gefunden wird [Bo79]. Es sorgt auch für eine Sicherung der Softwarequalität. Testen hat jedoch nur einen geringen Nutzen, wenn die Tests selbst eine schlechte Qualität haben und das Programm auch nach ausführlichem Testen noch viele Fehler aufweist. Um herauszufinden, wie gut ein Test ist, muss auch seine Qualität analysiert werden. Dazu eignet sich die in den 1970er Jahren entwickelte Technik des *Mutationstestens* [DLS78]. Hierbei werden Fehler in ein Programm eingebaut, welche die Tests entdecken müssen. Je mehr Fehler entdeckt werden, umso besser ist die Qualität der Tests.

Es existiert eine Menge unterschiedlicher Werkzeuge, mit denen Mutationstests automatisiert durchgeführt werden können [JH11]. Allerdings variiert die Leistungsfähigkeit dieser Werkzeuge. Dadurch erfordert die Entscheidung für ein geeignetes Werkzeug eine gründliche Abwägung der Vor- und Nachteile. Diese Arbeit beschäftigt sich mit dem Vergleich von ausgewählten Mutationstestwerkzeugen, kurz *MTW*, für die Programmiersprache Java und dem Testframework JUnit. Sie geht dabei der Frage nach, welches dieser MTW sich am besten für einen Einsatz in der Praxis eignet. Dazu erfolgt in Kapitel 2 zunächst eine kurze Einführung in den Bereich des Mutationstestens. In Kapitel 3 erfolgt die Vorstellung der untersuchten MTW und ihrer genutzten Ansätze. Kapitel 4 befasst sich mit der praktischen Umsetzung der Evaluation. Kapitel 5 zieht schließlich ein Fazit über die gewonnen Erkenntnisse.

[1]Universität Hildesheim, Abteilung Software Systems Engineering, Universitätsplatz 1, 31141 Hildesheim, gerlingl@uni-hildesheim.de

2 Mutationstests

Mutationstesten ist ein dynamisches, fehlerbasiertes Whitebox-Testverfahren, welches Ende der 1970er Jahre erstmals von DeMillo, Lipton und Sayward [DLS78] sowie Hamlet [Ha77] in wissenschaftlichen Publikationen vorgestellt wurde. Das Besondere an diesem Verfahren ist, dass es sich um eine Methode zur Qualitätsbestimmung einer Testmenge handelt. Getestet wird demnach nicht die Software selbst, sondern die dazugehörigen Tests. Ursprünglich war der Mutationstestansatz zur Überprüfung von Tests auf Komponentenebene gedacht [DLS78], mit der Zeit wurden zudem Anwendungs-möglichkeiten auf Integrations-, System- und Entwurfsebene geschaffen [JH11]. Auch die Anzahl der unterstützten Programmiersprachen hat sich seit den Anfängen deutlich erhöht. Zunächst gab es nur Mutationen für FORTRAN Programme, später kamen dann Sprachen wie Java, C oder SQL hinzu [JH11]. Aufgrund dieser Erweiterungen lassen sich Mutationstests heutzutage nach Offutt [Of11] folgendermaßen definieren: *„The use of well defined rules defined on syntactic descriptions to make systematic changes to the syntax or to objects developed from the syntax."*

Wie solch ein Mutationstest abläuft, sei an nachfolgendem Beispiel aus Tab. 1 erklärt. Zu Beginn steht die Eingabe eines Programms sowie der dazugehörigen Tests. Das Beispielprogramm erwartet die Eingabe einer Zahl. Ist die Zahl kleiner als 10, wird der Wert 0 zurückgegeben. Ist die Zahl größer oder gleich 10, wird der Wert der Zahl zurückgegeben. Der dazugehörige Test in der rechten Spalte von Tab. 1 überprüft dieses verhalten. Dazu wird in Zeile 1 überprüft, ob bei der Eingabe des Wertes 5 der Wert 0 zurückgegeben wird. In Zeile 2 wird geprüft, ob bei einer Eingabe von 15 auch der Wert 15 zurückgegeben wird. Beide Tests sind erfolgreich. Die erfolgreiche Durchführung der Tests ist eine notwendige Bedingung für fehlerbasiertes Testen [Mo90].

Programm	**Test**
1. int comp(int a){ 2. if (a < 10){ 3. return 0; 4. } else { 5. return a; 6. } 7. }	1.assertEquals(0,comp(5)); 2.assertEquals(15,comp(15));

Tab. 1: Beispielprogramm und Test

Als nächstes erfolgt die *Mutation* des Programms. Eine Mutation verändert immer nur einen kleinen Teil der Software, um realistische Fehler zu simulieren [DLS78]. Dies ist eine der Kernhypothesen, auf die sich Mutationstests stützen. Sie wird als *Competent Programmer Hypothesis* bezeichnet. Die Hypothese wurde folgendermaßen definiert: *„Programmers have one great advantage that is almost never exploited: they create programs that are close to being correct!"* [DLS78]. Sie besagt, dass kompetente Programmierer nur kleine Fehler machen, im Prinzip also nahezu korrekte Programme schreiben. Sie programmieren nicht zufällig, sondern versuchen ihr Programm so zu schreiben, dass es der Soll-Version möglichst nahe kommt.

Welche Arten von Veränderungen eingebaut werden, ist abhängig von den gewählten *Mutations-Operatoren* [JH11]. Allgemein definiert ein Mutations-Operator zwei Aspekte der Mutation: Was sind die Zielobjekte und welcher Art ist die Veränderung. Als Zielobjekte dienen beispielsweise Methodenaufrufe, mathematische Operatoren, Vergleichsoperatoren oder `return` Statements. Die Veränderung kann beispielsweise darin bestehen, Zielobjekte zu löschen, umzukehren oder ihren Wahrheitswert auf `true` zu setzen. Eine mögliche Mutation des Programms aus Tab. 1 wäre also beispielsweise, den Vergleichsoperator aus Zeile 2 umzukehren in `(a > 10)`. Neben diesen einfachen Mutationen, auch Mutationen auf Methodenebene genannt, gibt es noch weiterführende Konzepte wie beispielsweise objektorientierte Mutationen. Hier werden dann beispielsweise Sichtbarkeiten, Konstruktoren oder Schlüsselwörter wie `static`, `super` und `this` verändert. Außerdem gibt es noch die Möglichkeit, *Mutationen höherer Ordnung* durchzuführen. Das bedeutet, dass auch komplexere Mutationen erlaubt sind, um die Anzahl der insgesamt erzeugten Mutationen zu reduzieren [JH08].

Nach Erzeugung der mutierten Programmversionen, auch *Mutanten* genannt, werden die Testfälle für jeden Mutanten erneut ausgeführt. Daher ist es notwendig, dass die erzeugten Mutationen syntaktisch korrekt sind und die Software weiterhin kompilierbar ist [OU01]. Schlägt ein Testfall fehl, gilt der Mutant als *eliminiert*. Dies passiert auch bei der zuvor erwähnten Umkehrung des Vergleichsoperators des Beispielprogramms. Je mehr Mutanten eliminiert werden, desto besser sind die Tests. Ändert man den Vergleichsoperator des Beispielprogramms aus Tab. 1 beispielsweise in `(a <= 10)`, würden die Tests trotzdem noch erfolgreich durchlaufen. Diese Tests können also noch verbessert werden.

Allerdings gibt es auch Mutationen, die nicht automatisch entdeckt werden können, da sie semantisch äquivalent sind [BA82]. Hier spricht man dann von *äquivalenten Mutanten.* Diese Mutanten stellen eine Schwachstelle der Mutationstests dar, da sie die Bewertung der Tests verzerren können, wenn sie nicht als solche gekennzeichnet werden. Deswegen fließen die äquivalenten Mutanten bei der Berechnung des *Mutation Scores* mit ein. Dieser Wert steht am Ende des Mutationstestprozesses und wurde zuerst von Adrion, Branstad und Cherniavsky im Jahre 1982 eingeführt [ABC82]:

$$\text{Mutation Score} = \frac{|\text{eliminierte Mutanten}|}{|\text{erzeugte nicht-äquivalente Mutanten}|}$$

Der aus dieser Formel errechnete Wert gibt an, wie gut die getesteten Testfälle sind. Die Testfälle haben eine besonders hohe Qualität, wenn der Mutation Score einen Wert von 1 beziehungsweise 100% erreicht. Er berechnet sich aus der Anzahl an eliminierten Mutanten geteilt durch die Anzahl an nicht-äquivalenten Mutanten.

3 Vorstellung der Mutationstestwerkzeuge

MTW sind für die effiziente Anwendung von Mutationstests nahezu unersetzbar, da eine

manuelle Ausführung einen hohen Aufwand erfordert. Doch die Entscheidung für eines dieser MTW erfordert die Einbeziehung verschiedener Faktoren. Als besonders wichtig wurden dabei die Faktoren Verfügbarkeit, unterstützte Java-Version, Plattformunabhängigkeit, sowie Status der Entwicklung gewertet. Daher werden in dieser Arbeit nur die MTW betrachtet, die frei verfügbar zum Download stehen, mindestens Java Version 5 unterstützen, plattformunabhängig funktionieren und innerhalb der letzten 3 Jahre weiterentwickelt wurden. Dieses Kapitel stellt die ausgewählten MTW und ihre Eigenschaften vor.

3.1 Übersicht der Werkzeuge

MuJava ist das älteste MTW für Java, welches noch aktiv weiterentwickelt wird. Seine Ursprünge hat es in dem 2002 von Chevalley und Thévenod-Fosse entwickeltem Werkzeug JavaMut [CT03]. Die erste Version von MuJava wurde 2003 noch unter dem Namen JMutation von Offutt, Ma, und Kwon veröffentlicht [OMK04]. Diese Arbeit untersucht Version 4, welche im Juni 2013 veröffentlicht wurde. Beispiele und genauere Erläuterungen der verwendeten Mutationen von MuJava finden sich in [MO11] und [MO14].

Das MTW *Jumble* wurde erstmals 2007 von Irvine et al. veröffentlicht [Ir07]. Zuvor wurde es von 2003-2006 als Werkzeug der Firma Reel Two entwickelt. Es wurde dazu entworfen, in einer industriellen Umgebung mit großen Programmen zu arbeiten. In dieser Arbeit wird Version 1.2.0 vom 22.04.2013 untersucht. Eine ausführliche Erklärung der Mutations-Operatoren findet sich auf [Ju07].

Das von Henry Coles entwickelte MTW *Pitest* wurde zum ersten Mal im Dezember 2011 veröffentlicht [Co14a]. Es basierte ursprünglich auf Jumble, benutzt aber mittlerweile eigene Ansätze zur Mutanten-Erzeugung. Die betrachtete Version ist Version 1.0.0 vom 18. Mai 2014. Eine Erklärung zu Pitests Mutations-Operatoren findet sich auf [Co14b].

Judy ist ein von Lech Madeyski und Norbert Radyk im Jahr 2011 veröffentlichtes MTW [MR10]. In dieser Arbeit wird Version 2.1.0 vom 26. Januar 2014 untersucht. Erläuterungen und Beispiele zu den Mutations-Operatoren von Judy finden sich in [MR10].

Die erste Version des MTW *Major* wurde im Januar 2014 von René Just veröffentlicht [Ju14]. Somit ist Major das jüngste der betrachteten MTW. Diese Arbeit befasst sich mit der aktuellen Version 1.1.2, welche am 27. Juni 2014 zum Download bereitgestellt wurde. Eine genaue Beschreibung von Majors Mutationen findet sich in [Ma14].

3.2 Gegenüberstellung der Eigenschaften

Jedes der MTW bietet dem Benutzer die Möglichkeit, die angewendeten Mutations-Operatoren auszuwählen, siehe Tab. 2. Dieses Vorgehen nennt sich selektive Mutation.

Es dient auch dazu, die Erzeugung äquivalenter Mutanten zu verringern. Dies geschieht, indem nur die Operatoren ausgewählt werden, die statistisch gesehen sehr wenig äquivalente Mutanten erzeugen. Allerdings lässt sich dadurch die Erzeugung nur indirekt und bis zu einem gewissen Maße verringern, mehr dazu findet sich auch in [MB99]. Methoden zur Identifizierung äquivalenter Mutanten werden nicht angewendet. Weiterhin fällt der Umfang der Selektionsmöglichkeiten der einzelnen MTW recht unterschiedlich aus. Während MuJava 41 Mutations-Operatoren zur Auswahl stellt, sind es bei Jumble nur Sieben.

Name	Selektive Mutation	Mutationen Höherer Ordnung	Objektorientierte Mutationen
MuJava	Ja	Nein	Ja
Jumble	Ja	Nein	Nein
Pitest	Ja	Nein	Nein
Judy	Ja	Ja	Ja
Major	Ja	Nein	Nein

Tab. 2: Vergleich der Mutationsarten

MuJava und Judy nutzten im Gegensatz zu den anderen MTW sowohl Mutations-Operatoren auf Methodenebene als auch objektorientierte Operatoren, wie in Tab. 2 zu sehen ist. Major bietet als einziges MTW die Möglichkeit, Operatoren nicht nur ein- und auszuschalten, sondern auch ihre Funktionsweise zu konfigurieren. Dafür nutzt Judy als einziges MTW noch zusätzlich die Technik der Mutationen höherer Ordnung.

Um die Testausführung zu optimieren, nutzt MuJava als einziges MTW keine Techniken, siehe Tab. 3. Jumble, Pitest und Major hingegen kombinieren drei verschiedene Methoden zur Priorisierung. Auch wenn sich diese Methoden innerhalb der MTW unterscheiden, so geht es doch im Kern darum, nur die Tests mit der kürzesten Dauer auszuführen, die zu der aktuell betrachteten Mutation gehören und die im Vergleich zum letzten Testlauf ein anderes Ergebnis erwarten lassen. Weiterhin nutzen Jumble, Judy und Major noch jeweils eine andere Technik zur Verbesserung der Ausführungszeiten. Zudem nutzten alle MTW eine Begrenzung der Ausführungsdauer der Tests, um Endlosschleifen zu vermeiden.

Name	Testausführungsoptimierung
MuJava	Keine
Jumble	Priorisierung nach Dauer, Änderung und Zugehörigkeit, Parallele JVM
Pitest	Priorisierung nach Dauer, Änderung und Zugehörigkeit
Judy	Pointcut and Advice
Major	Priorisierung nach Dauer, Änderung und Zugehörigkeit, Weak Mutation

Tab. 3: Testausführungsoptimierung der MTW

Die nachfolgende Tab. 4 enthält die Eigenschaften der MTW, welche für die technische Integration relevant sind. MuJava, Jumble und Judy unterstützen Java Version 6, Major

unterstützt Version 7 und Pitest Version 8. Alle MTW unterstützen JUnit 4. Sämtliche MTW außer MuJava können über Ant betrieben werden. Für Jumble und Pitest gibt es zusätzlich noch ein Eclipse Plugin, welches aber nicht alle Konfigurationsmöglichkeiten des Kommandozeilen-Betriebs bietet. Weiterhin unterstützen beide noch einige Mock-Versionen und Pitest bietet zusätzlich noch Unterstützung für Maven, Gradle, TestNG, Sonar und Intellij. Da es sich um MTW für Java handelt, sind sie prinzipiell nicht an ein Betriebssystem gebunden. Allerdings funktioniert die Installation von Judy über Shell bzw. Bat Skripte, wodurch eine einfache Installation nur unter Windows und Linux möglich ist.

Name	Java Version	JUnit Version	Toolunterstützung	Betriebssystem
MuJava	6	4	-	Plattformunabhängig
Jumble	6	4	Ant, Eclipse, Mock	Plattformunabhängig
Pitest	8	4	Ant, Maven, Gradle, TestNG, Eclipse, Mock, Sonar, Intellij	Plattformunabhängig
Judy	6	4	Ant	Windows, Linux
Major	7	4	Ant	Plattformunabhängig

Tab. 4: Technische Eigenschaften der MTW

4 Leistungsvergleich der Mutationstestwerkzeuge

Mutationstests finden nur selten eine Anwendung in der Praxis [JH11]. Dies hängt unter anderem mit ihrem relativ hohen Ressourcenverbrauch zusammen. Daher wird in diesem Kapitel die Frage beantwortet, wie gut die getesteten MTW mit den Performanzproblemen umgehen. Weiterhin stellt sich auch die Frage, ob die MTW überhaupt für einen Praxiseinsatz in Wirtschaft oder Forschung geeignet sind. Hohe Kosten sind kein Argument, wenn sich der Aufwand lohnt. Daher liegt die Vermutung nahe, dass es noch weitere Probleme bei der Anwendung gibt. Deswegen wird in diesem Kapitel zusätzlich auch die Bedienbarkeit und Aussagekraft der MTW überprüft.

4.1 Beschreibung der Testumgebung

Inhalt dieser Arbeit ist der Vergleich von MTW für die Programmiersprache Java. Um die Aktualität der Ergebnisse zu gewährleisten, wurde daher mit der Java Laufzeitumgebung Version 7.65 getestet. Als Systemumgebung wurde eine virtuelle Maschine mit Ubuntu 14.04 LTS verwendet. Das System wurde mit 2 GB RAM und einem Dualcore Prozessor mit 3,3 GHz aufgesetzt. Zur Ausführung der Evaluation wurden das Linux Terminal, Eclipse Luna Version 4.4.0 und JUnit Version 4.11 verwendet. Leistungsdaten wurden mit der Java VisualVM Version 1.3.3 erhoben.

Nicht jedes MTW ist eigenständig dazu in der Lage, in einem Testlauf mehrere Klassen zu mutieren. Um die Vergleichbarkeit der Ergebnisse zu gewährleisten, wurde daher als

Testobjekt ein sehr simples Programm verwendet. Dieses Programm dient zur Klassifizierung von Dreiecken und wurde ursprünglich von Kevin Jalbert als Testbeispiel für das MTW Javalanche geschrieben [Ja12]. Es besteht aus einer Klasse und zwölf JUnit-Tests.

4.2 Vorstellung der Testkriterien

Die hier vorgestellten Kriterien finden in Abschnitt 4.4 ihre Anwendung. Dabei enthält Tab. 5 die nachfolgend erläuterten Leistungskriterien. Das erste Kriterium ist der *Mutation Score*. Er gibt Auskunft über die Qualität der getesteten Testmenge. Außerdem dient er als Effektivitätskriterium zur Bewertung der MTW, da die Berechnung dieses Wertes ein Hauptanliegen der Mutationstests ist. Ein MTW funktioniert nur effektiv, wenn es einen sinnvollen Mutation Score berechnen kann. Damit verbunden ist die *Anzahl der erzeugten Mutanten*. Ist sie zu gering, hat der erzeugte Mutation Score keine Aussagekraft. Ein weiteres Effektivitätskriterium ist die *Anzahl der ausgeführten JUnit Tests* zur Eliminierung der Mutanten. Auch hier bedeutet ein zu geringer Wert einen Verlust an Effektivität. Ein zu hoher Wert hingegen kann ein Anzeichen für unnötig ausgeführte Tests sein.

Als Effizienzkriterium dient unter anderem der *maximale Speicherverbrauch* während eines Durchlaufs. Der Verbrauch war bei jedem MTW während des Testlaufs konstant. Weiterhin wurde die durchschnittliche *CPU-Auslastung* pro Testlauf gemessen. In Klammern steht die durchschnittliche Abweichung von diesem Wert. Diese Abweichung zeigt an, ob es starke Schwankungen während eines Testlaufs gibt. Die letzten beiden Effizienzkriterien sind die *Anzahl der erzeugten Mutanten pro Sekunde* und der *ausgeführten Tests pro Sekunde*. Hierbei wird noch einmal besonderes Augenmerk auf die effiziente Erzeugung der Mutanten beziehungsweise effiziente Durchführung der Tests zur Eliminierung der Mutanten gelegt.

Neben den objektiven Effektivitäts- und Effizienzbewertungen gibt es auch subjektive Bewertungen hinsichtlich der Bedienbarkeit der MTW. Diese werden in Tab. 6 angewendet und im nachfolgenden Absatz erläutert. Da sich die Bedienbarkeit eines MTW nur schwer objektiv messen lässt, wird jedes Kriterium stattdessen mit einer Note von 1 (*sehr gut)* bis 5 (*mangelhaft)* bewertet. Diese Note fasst die Erfüllung der Teilaspekte in einem leicht zu interpretierenden Wert zusammen.

Zum einen wurden *Installation*, *Mutanten-Erzeugung* und *Testausführung* hinsichtlich der Schwierigkeit der Ausführung dieser Schritte bewertet. Hier haben die MTW besonders gut abgeschnitten, bei denen sich diese Aktionen mit wenig Aufwand zügig erledigen lassen. *Dokumentation* und *Ergebnisdarstellung* wurden hinsichtlich des enthaltenen Umfangs bewertet und der in diesem Umfang enthaltenen Informationen. Hier steht Qualität über Quantität. Ähnlich verhält es sich auch mit der *Konfigurierbarkeit*. Die Anzahl der Möglichkeiten spielt eine Rolle, allerdings ist auch die Umsetzung nicht zu vernachlässigen. Falls die Konfiguration sehr umständlich ist, oder angegebene Möglichkeiten nicht umsetzbar sind, führt dies zu einer Abwertung.

4.3 Vorgehensbeschreibung

Zu Beginn der Evaluation erfolgte einmalig das Aufsetzen einer neuen virtuellen Maschine mit der in Abschnitt 4.1 beschriebenen Konfiguration. Innerhalb dieser Testumgebung wurde dann jeweils die neueste Version des MTW heruntergeladen. Auf Grundlage der verfügbaren Dokumentation wurde jedes MTW genau nach Anweisung installiert. Damit keine Seiteneffekte auftreten, wurden sämtliche Änderungen am System nach Erhebung der Testdaten rückgängig gemacht. Die Ausführung der Tests wurde 20-mal wiederholt. Die letztendlichen Ergebnisdaten wurden mithilfe des Median ermittelt, um Ausreißern keine zu große Gewichtung zu geben. Die MTW wurden mit den empfohlenen Standardeinstellungen getestet.

4.4 Darstellung der Ergebnisse

In Tab. 5 findet sich eine Übersicht über die einzelnen Bewertungen. Ein Fragezeichen bei der Bewertung bedeutet, dass dieses Kriterium nicht gemessen werden konnte beziehungsweise dazu keine Angaben gemacht wurden.

Betrachtet man den ausgegebenen Mutation Score, so fällt auf, dass dieser Wert starken Schwankungen unterworfen ist. Da immer die gleiche Testmenge bewertet wurde, liegt die Ursache dieser Schwankungen in der Methodik der MTW. Da nur MuJava und Major genaue Angaben über die erzeugten Mutationen machen, lässt sich nicht eindeutig feststellen, warum der Mutation Score so unterschiedlich ausfällt. Weiterhin scheint der berechnete Wert von MuJava nicht korrekt zu sein, da die erzeugten Mutanten zumindest teilweise von den Tests hätten entdeckt werden müssen. Die Gründe dafür sind nicht nachvollziehbar. Bei den anderen MTW hängen die unterschiedlichen Ergebnisse vermutlich mit der Anzahl der äquivalenten Mutanten und den genutzten Mutations-Operatoren zusammen. Dieselbe Argumentation betrifft auch die Anzahl der eliminierten Mutanten.

	MuJava	**Jumble**	**Pitest**	**Judy**	**Major**
Mutation Score	0%	97%	82%	100%	67%
Erzeugte Mutanten	432	39	44	121	79
Eliminierte Mutanten	0	38	36	121	53
Ausführungsdauer	36s	1,7 s	4,8 s	3,1 s	3,5 s
Ausgeführte Tests	?	?	116	662	641
Mutanten / Sekunde	12	22,9	9,2	39	22,6
Tests / Sekunde	?	?	24,2	213,5	183,1
Speicherverbrauch	232MB	5 MB	19 MB	95 MB	?
CPU-Auslastung	66% (± 27%)	2,6% (± 0,2%)	27% (± 3,1%)	0,01% (± 0%)	?

Tab. 5: Ergebnisse der Leistungstests

Die Anzahl der erzeugten Mutanten hängt mit den anwendbaren Mutations-Operatoren

zusammen. Je mehr Operatoren aktiv sind, desto mehr Mutanten können auch erzeugt werden. Eine höhere Anzahl erzeugter Mutanten ist einerseits positiv, da somit die Aussagekraft des Mutation Scores steigt. Andererseits bedeuten mehr Mutanten aber auch einen höheren Ressourcenbedarf. Deswegen ist es sinnvoll, auch die erzeugten Mutanten pro Sekunde zu betrachten. Hier fällt auf, dass Judy die Mutanten mit am effizientesten erzeugt. Klarer Verlierer hingegen ist das MTW Pitest, welches nicht nur vergleichsweise wenig Mutanten erzeugt, sondern dies auch am ineffizientesten tut.

Bei der Anzahl der durchgeführten Tests zur Eliminierung der Mutanten liegen Judy und Major mit weitem Abstand vorne, wobei hier auch nicht jedes MTW Daten dazu ausgegeben hat. Die unterschiedlichen Ergebnisse lassen sich durch die Anwendung verschiedener Optimierungsmethoden bei der Testausführung erklären, siehe Tab. 3.

Judy und Major nutzen beide Techniken zur Verbesserung der Effizienz, welches sich auch in der großen Anzahl der ausgeführten Tests pro Sekunde erkennen lässt. Pitest hingegen arbeitet mit Selektion der durchzuführenden Tests, wodurch die vergleichsweise geringe Anzahl an durchgeführten Tests zustande kommt. Allerdings ist auch Pitests Anzahl der ausgeführten Tests pro Sekunde im Vergleich sehr gering.

Bei den Daten zum Leistungsverbrauch konnte Major nicht gemessen werden, da es mit einer eigenen VM arbeitet. Judy überzeugt durch eine extrem niedrige CPU-Auslastung, hat dafür aber einen hohen Verbrauch an Arbeitsspeicher. Bei der Ausführungsdauer liegen Judy, Major und Pitest im Mittelfeld. Am schlechtesten schneidet MuJava ab, da es mit Abstand am längsten braucht und die höchste CPU-Auslastung hervorruft sowie am meisten Arbeitsspeicher benötigt. Jumble hingegen hat die kürzeste Ausführungsdauer, verbraucht sehr wenig RAM und benötigt auch nur eine geringe CPU-Auslastung.

Nachfolgend werden die Ergebnisse der Bedienbarkeitsbewertung diskutiert. Die einzelnen Bewertungen sind dabei in Tab. 6 zusammengestellt. Es sei anzumerken, dass die Benotung subjektiver Wahrnehmung entspricht und somit von Betrachter zu Betrachter abweichen kann.

	MuJava	**Jumble**	**Pitest**	**Judy**	**Major**
Installation	3	1	1	1	1
Dokumentation	3	2	3	3	5
Mutanten-Erzeugung	2	1	1	2	2
Testausführung	2	1	1	2	2
Ergebnisdarstellung	2	3	2	2	2
Konfigurierbarkeit	5	2	2	3	1
Automatisierbarkeit	4	1	1	1	1

Tab. 6: Ergebnisse der Bedienbarkeitstests

Bei der Bewertung der Installation gibt es nur zwei verschiedene Ausprägungen. Zum einen haben die MTW Jumble, Pitest, Judy und Major alle eine sehr gute Bewertung bekommen, da die Installation nur aus dem Herunterladen und eventuellen Entpacken

der MTW besteht. Einzig MuJava hat schlechter abgeschnitten, da hier eine Anpassung der Umgebungsvariablen und das Ändern einer Konfigurationsdatei notwendig waren.

Die Dokumentation zu den MTW war bestenfalls gut, im Falle von Major sogar nur mangelhaft. Letzteres kommt daher, dass Major zwar eine sehr umfangreiche Dokumentation besitzt, die wirklich wichtigen Informationen aber nicht enthalten sind. Die beste Dokumentation gehört zum MTW Jumble. Die vorhandenen Beschreibungen sind zwar kurz gehalten, dafür sind trotzdem alle wünschenswerten Informationen enthalten und übersichtlich dargestellt. MuJava, Pitest und Judy besitzen nur befriedigende Dokumentationen, da teilweise Lücken vorhanden oder fehlerhafte Informationen angegeben sind. Die Auswirkungen dieser Fehler sind allerdings nur minimal und führen daher nur zu einer geringen Abwertung.

Was Mutanten-Erzeugung und Testausführung angeht, haben alle MTW gut bis sehr gut abgeschnitten. Jumble und Pitest haben den Vorteil, dass durch ihr Eclipse Plugin die Ausführung dieser Aktionen sehr simpel ist. Auf Knopfdruck lassen sich Mutanten-Erzeugung und Testausführung in einem Schritt starten. Auch bei MuJava, Judy und Major lassen sich diese Schritte gut und einfach durchführen. MuJava verwendet dabei eine grafische Oberfläche, die über die Kommandozeile aufgerufen wird. Judy und Major benötigen nur die Eingabe über die Kommandozeile mit der Übergabe einiger Parameter. Das ist zwar nicht ganz so intuitiv wie die Ausführung direkt in Eclipse, aber dennoch ohne großen Aufwand zu bewerkstelligen.

Bei der Darstellung der Ergebnisse hat Jumble die schlechteste Note bekommen. Es werden zwar solch wesentliche Angaben wie Mutation Score und überlebende Mutanten ausgegeben, Angaben über die erzeugten Mutanten oder nützliche Statusinformationen fehlen hingegen. MuJava, Pitest, Judy und Major haben eine gute Darstellung der Ergebnisse. Hier fehlten nur die Angaben über die erzeugten Mutationen beziehungsweise bei Major und MuJava eine übersichtliche Darstellung. Bei Major werden zwar die Mutationen genannt, allerdings benötigt man drei verschiedene Logdateien, um mit den Informationen auch etwas anfangen zu können. Bei MuJava werden die Logausgaben und die grafische Oberfläche zum Nachvollziehen der Detailinformationen benötigt.

Was die Konfigurierbarkeit der MTW angeht, haben Jumble, Pitest und Judy alle ähnliche Möglichkeiten zur Verfügung wie beispielsweise Anpassung der Mutations-Operatoren oder der Testklassen. Judy hat allerdings eine Abwertung bekommen, da einige der Konfigurationsmöglichkeiten nicht funktioniert haben, wie beispielsweise die Ausgabe von Debug-Informationen oder der erzeugten Mutationen. Klarer Sieger ist Major, da durch seine eigene domänenspezifische Sprache eine Vielzahl an Konfigurationsmöglichkeiten der Mutationen gegeben ist. Der Verlierer in Sachen Bedienbarkeit ist MuJava, da hier außer der Selektion der Mutations-Operatoren keine Konfigurationsmöglichkeiten gegeben sind.

Das letzte Bedienbarkeitskriterium ist die Automatisierbarkeit des Testprozesses. Auf diesem Punkt liegt noch einmal besonderes Augenmerk, da die benötigte menschliche

Beteiligung ein großer Kritikpunkt der Mutationstests ist. Hier ist erfreulicherweise festzustellen, dass alle MTW außer MuJava komplett automatisierbar sind. MuJava hat eine ausreichende Bewertung bekommen, da es Ant nicht unterstützt und die Bedienung über eine grafische Oberfläche erfolgt.

5 Fazit

Die Ergebnisse der Evaluation haben gezeigt, dass MTW durchaus eine Existenzberechtigung in der modernen Softwareentwicklung haben. Ihre Anwendung ist voll automatisierbar und auch die Performanzprobleme sind mittlerweile kein großes Problem mehr. Allerdings lässt die Bedienbarkeit teilweise noch stark zu wünschen übrig und auch die Nachvollziehbarkeit der ausgegebenen Ergebnisse ist nicht immer gewährleistet. Daher können von den getesteten MTW nur Jumble und Pitest für einen praktischen Einsatz empfohlen werden. Major eignet sich nur, falls man schon Erfahrung mit Mutationstests gesammelt hat und ein hoch-konfigurierbares MTW benötigt. Von MuJava und Judy ist wegen der mangelnden Nachvollziehbarkeit des berechneten Mutation Scores abzuraten.

Obwohl es ständig neue Entwicklungen im Bereich der Mutationstests gibt und die getesteten MTW ihre Tauglichkeit bewiesen haben, ist die praktische Anwendung immer noch wenig verbreitet. Dies ändert sich hoffentlich in den nächsten Jahren, wenn die Vorurteile gegenüber den MTW beseitigt werden können. Dazu wäre es natürlich auch hilfreich, wenn die Bedienbarkeit weiter verbessert würde. Vielleicht wäre es hier sinnvoll, bei der Entwicklung neuer MTW einen größeren Fokus auf ihre Praxistauglichkeit zu legen.

Literaturverzeichnis

[ABC82] Adrion, W. R.; Branstad, M. A.; Cherniavsky, J. C.: Validation, verification, and testing of computer software. ACM Computing Surveys (CSUR), Volume 14, Nr. 2, S. 159–192, 1982.

[BA82] Budd, T. A.; Angluin, D.: Two Notions of Correctness and Their Relation to Testing. Acta Informatica, Volume 18, Nr. 1, S. 31–45, 1982.

[Bo79] Boehm, B. W.: Guidelines for verifying and validating software requirements and design specifications. In Proc. EURO IFIP'79, S. 711-719, London, Sep. 1979.

[Co14a] Coles, H.: Pitest., http://pitest.org, Stand: August 2014.

[Co14b] Coles, H.: Pitest Mutators., http://pitest.org/quickstart/mutators/, Stand: August 2014.

[CT03] Chevalley, P.; Thévenod-Fosse, P.: A mutation analysis tool for Java programs. Internal Journal on Software Tools for Technology Transfer, Volume 5, Nr. 1, S. 90-103, 2003.

[DLS78] DeMillo, R. A.; Lipton, R. J.; Sayward, F. G.: Hints on test data selection: Help for the practicing programmer. IEEE Computer 11 (4), S. 34–41, 1978.

[Ha77] Hamlet, R. G.: Testing Programs with the Aid of a Compiler. IEEE Transactions on Software Engineering, Volume 3, Nr. 4, S. 279-290, 1977.

[Ir07] Irvine, S. A.; Pavlinic, T.; Trigg, L.; Cleary, J. G.; Inglis, S.; Utting, M.: Jumble Java byte code to measure the effectiveness of unit tests. In Testing: Academic and Industrial Conference Practice and Research Techniques-MUTATION, S. 169–175, 2007.

[Ja12] Jalbert, K.: Triangle Example., https://github.com/david-schuler/javalanche/tree/master /examples/triangleJunit4, Stand: 30. Januar 2012.

[JH08] Y. Jia and M. Harman, "Constructing subtle faults using higher order mutation testing," in Eighth IEEE International Working Conference on Source Code Analysis and Manipulation, S. 249–258, 2008.

[JH11] Jia, Y.; Harman, M.: An Analysis and Survey of the Development of Mutation Testing. IEEE Transactions on Software Engineering, Volume 37, Nr. 5, S. 649–678, Sep. 2011.

[Ju07] Jumble Mutations., http://jumble.source-forge.net/mutations.html, Stand: 6. Juni 2007.

[Ju14] Just, R.: The Major mutation framework: Efficient and scalable mutation analysis for Java. In Proceedings of the International Symposium on Software Testing and Analysis (ISSTA), 2014.

[Ma14] Major Mutation Framework., http://mutation-testing.org/doc/, Stand: Juni 2014.

[MB99] Mresa, E. S.; Bottaci, L.: Efficiency of mutation operators and selective mutation strategies: An empirical study. Software Testing Verification and Reliability, Volume 9, Nr. 4, S. 205–232, 1999.

[MO11] Ma, Y.-S.; Offutt, J.: Description of class mutation operators for java. 2011.

[MO14] Ma, Y.-S.; Offutt, J.: Description of method-level mutation operators for java. 2014.

[Mo90] Morell, L. J.: A theory of fault-based testing. IEEE Transactions on Software Engineering, Volume 16, Nr. 8, S. 844–857, 1990.

[MR10] Madeyski, L.; Radyk, N.: Judy – a mutation testing tool for Java. IET Software, Volume 4, Nr. 1, S. 32, 2010.

[Of11] Offutt, J.: A mutation carol: Past, present and future. Information and Software Technology, Volume 53, Nr. 10, S. 1098–1107, 2011.

[OMK04] Offutt, J.; Ma, Y.-S.; Kwon, Y.-R.: An experimental mutation system for Java. ACM SIGSOFT Software Engineering Notes, Volume 29, Nr. 5, S. 1–4, 2004.

[OU01] Offutt, J.; Untch, R. H.: Mutation 2000: Uniting the orthogonal. Mutation testing for the new century, Springer US, S. 34-44, 2001.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Patterns im Umfeld prozessgetriebener Anwendungen zur Dokumentenverarbeitung mit BPMN

Tim Mühlsteff [1] und Andreas Schäfer [2]

Abstract: An prozessgetriebene Anwendungen werden hohe Anforderungen bezüglich der Wartbarkeit und Bedienbarkeit gestellt. Standardpatterns, angepasst an die Bedürfnisse einer prozessgetriebenen Anwendung, erleichtern es Entwicklern diese Anforderungen zu erfüllen. Konkret wird in diesem Paper gezeigt, dass sich das MVP-Pattern mithilfe eines EventBus-Systems unkompliziert mit Vaadin einsetzen lässt. Über eine Factory in Verbindung mit einer sicher gestalteten Verwendung von Reflections können zusätzliche User Tasks flexibel in der UI verwendet werden, ohne großen Bearbeitungsaufwand im Quellcode zu verursachen. Es wird ein dreiteiliges Design-Pattern vorgestellt, das dem Benutzer alle Informationen übersichtlich darstellt. Genehmigungsprozesse, die in der Modellierung ein Statuskonzept integrieren, ermöglichen eine gute Nutzung von Prozessen über Abteilungsgrenzen hinweg. Weiter ermöglicht BPMN die Prozess-Einzelverarbeitung eines Dokumentes mithilfe eines weiteren Prozesses, um eine Mengenverarbeitung mehrerer Dokumente zu ergänzen.

Keywords: Prozess, User Interface, Pattern, Factory, MVC, MVP, BPMN, Camunda, Vaadin

1 Einleitung

In prozessgetriebenen Anwendungen kann es vorkommen, dass die Modellierung des Geschäftsprozesses sehr komplex wird. Dies kann insbesondere dann der Fall sein, wenn Prozessautomatisierung zum Einsatz kommt. Daher sollte die Systemarchitektur möglichst genau spezifiziert werden. Das führt dazu, dass der Quellcode dieser Anwendung besser strukturiert werden kann und damit wartbarer ist. Zudem sollte auch die Konzeption der User Interface (UI) Komponente bedacht werden. Sie sollte für den Nutzer möglichst leicht und intuitiv bedienbar sein sowie alle relevanten Informationen übersichtlich darstellen. Mit BPMN (Business Process Model and Notation) wird eine Grundlage für die Modellierung von Geschäftsprozessen geboten. Aktuell steht sie in der Version 2.0 zur Verfügung [OM11]. Die Notation ist so strukturiert, dass sich mit Hilfe nur weniger Komponenten Prozesse sowohl aus fachlicher als auch aus technischer Sicht gut abbilden lassen. Zentrale Bestandteile sind dabei vor allem Aktivitäten, Ereignisse sowie Gateways. [FR12, S.21]

Das Open Source Projekt Camunda stellt eine eigene BPM Plattform zur Verfügung, die eine Ausführung automatisierter Prozesse ermöglicht. Dies erfolgt mittels der Camunda Process Engine, die das technische Prozessmodell steuert. [FR12, S.6f]

[1] Fachhochschule Münster, Corrensstraße 25, 48149 Münster, muehlsteff@fh-muenster.de
[2] Fachhochschule Münster, Corrensstraße 25, 48149 Münster, andreas.schaefer@fh-muenster.de

Anhand von BPMN und der darauf aufbauenden Process Engine von Camunda werden im Folgenden Patterns beschrieben, die vor allem im Zusammenhang mit prozessgesteuerten sowie aufgabenorientierten Dokumentenverarbeitungsprozessen eingesetzt werden können.

Im Rahmen dieses Papers werden unterschiedliche Patterns zur Unterstützung automatisierter Prozesse vorgestellt und bewertet. So wird mit den User-Interface Patterns zunächst auf die Trennung von UI- und Geschäftslogik unter Verwendung des Model-View-Presenter (MVP) Patterns näher eingegangen. Dabei wird die Problematik betrachtet, dass der Presenter stets die Steuerung der Benutzeroberfläche übernehmen sollte. Es werden Wege aufgezeigt, wie Steuerungslogik aus der View in den Presenter verschoben werden kann. Ein weiterer Aspekt im Bereich UI ist die Gestaltung und der Aufbau der Verarbeitungsmaske und was in Verbindung mit aufgabenorientierten Prozessen besonders zu beachten ist. Zusätzlich wird auf das technische Mapping zwischen User Task und UI-Maske eingegangen und ein Weg aufgezeigt, der die möglichen Probleme bei der Verwendung von Reflections vermeidet. Im zweiten Teil des Papers stehen Process Patterns im Fokus. Zunächst werden unterschiedliche Möglichkeiten zur Erfassung von Zuständen bzw. Status eines Dokumentes in einem Geschäftsprozess erläutert. Im Weiteren wird darauf eingegangen, wie in einer BPMN-Anwendung Daten zu mehreren Einzeldokumente mit einem bestimmten Zustand in einer Maske erfasst und gespeichert werden können.

2 User-Interface Patterns

In diesem Kapitel geht es um die Vorstellung UI-orientierter Patterns, insbesondere der Anpassung des MVP-Patterns zur Trennung von Darstellungs- und Anwendungslogik, sowie um die wiederverwendbare Strukturierung von UI-Masken.

2.1 Anwendung des Model View Presenter Patterns mit Vaadin

Eine Anwendung soll auf verschiedenen Plattformen und Gerätearten angeboten werden. Dazu sind unterschiedliche Darstellungsformen notwendig. Zudem soll die Anwendungslogik, die im Kontext einer prozessgetriebenen Anwendung vor allem aus der Kommunikation mit der Process Engine und anderen externen Systemen besteht, nicht vermischt werden. Ein bekanntes Entwurfsmuster, das dieses Problem lösen soll, ist das Model-View-Controller (MVC) Pattern und dessen Weiterentwicklung, das Model-View-Presenter (MVP) Pattern. Im MVC Pattern kennt die View das Modell und den Controller, sie nutzt die Objekte aus dem Modell, um die Daten darzustellen. Zusätzlich bearbeitet sie auch Benutzerinteraktionen. Im MVP Pattern kennt die View das Modell nicht. Ebenfalls hat sie keinen direkten Zugriff auf den Presenter und enthält keine Logik zur Steuerung der UI (siehe Abb. 1). Es ist alleinige Aufgabe des Presenters auf das Modell zuzugreifen und die View zu manipulieren. [Fo06]

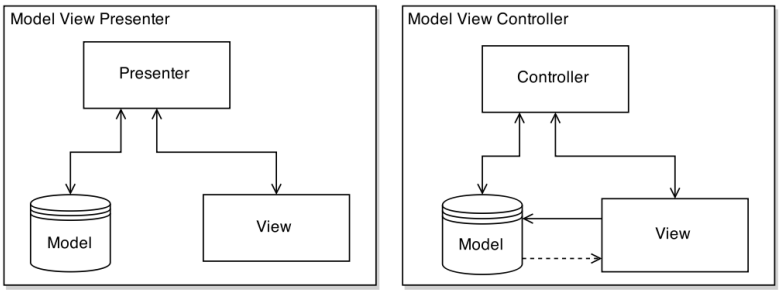


Abb. 1: Model View Presenter im Vergleich zum Model View Controller Pattern

Es besteht die Anforderung, dass die Anwendung über ein Web-Interface erreichbar sein soll. Für die Entwicklung einer Web-Oberfläche unter Java werden verschiedene Frameworks angeboten, die unterschiedliche Schwerpunkte setzen. Das Framework Vaadin bietet den Vorteil, eine Web-UI ausschließlich mit Java-Quellcode zu entwickeln. Es ist optional möglich eigene HTML-Templates zu verwenden. [VA15]

Das Framework Vaadin stellt Entwicklern Listenern auf Basis des Obersever-Patterns bereit, um auf Benutzerinteraktionen reagieren zu können. Es handelt sich dabei um eine Push-Variante in der eine definierte Funktion aufgerufen wird: [GM14, Kap. 5.14]

```
button.addClickListener(new Button.ClickListener() {
   public void buttonClick(ClickEvent event) { ... }
});
```

Im obigen Beispiel wird ein neuer Listener instanziiert und der Button-Variable zugewiesen. Dieser Codeausschnitt ist in der View platziert, die alle UI-Komponenten als Java Objekte beinhaltet. Nach dem MVP-Pattern darf an dieser Stelle keine Steuerungslogik implementiert werden, sondern der Presenter muss das Ereignis verarbeiten. Über EventBus-Systeme kann der Presenter informiert werden. Mit ihnen können aus allen Objekten heraus Ereignisse in einen Bus gesendet werden. Methoden, die als Subscriber beim EventBus registriert sind, werden bei Vorlage eines passenden Ereignisses aufgerufen. Damit kann folgendes Szenario realisiert werden: Aus der Klick-Methode der Schaltfläche kann eine Publish-Nachricht an den EventBus gesendet werden. Im Presenter wird eine Subscriber-Methode implementiert, die beim Anklicken des Buttons ausgeführt werden soll. Ein Open-Source Eventbussystem ist in den weit verbreiteten Google Guava Bibliotheken für Java enthalten. Die Identifikation des konkreten Eventstyps, z.B. *Benutzter klickt auf Schaltfläche A*, erfolgt über einen Java-Typ. Die instanziierten Objekte speichern ggf. Informationen zum Event in Variablen und stellen diese über Getter-Methoden dem Empfänger zur Verfügung. [WL12]

Auf der Webseite von Vaadin wird eine Alternative vorgestellt, die ausschließlich auf Vaadin-Komponenten basiert. Als Beispiel wird eine Taschenrechner-Anwendung verwendet (siehe Abb. 2) [GM13].

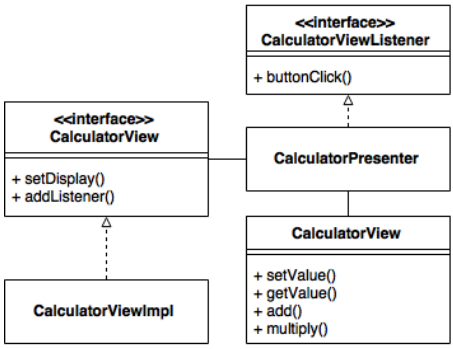


Abb. 2: UML-Klassendiagramm aus dem MVP-Beispiel von Vaadin [GM13]

In diesem Fall ist die View eine CustomComponent, die der Entwickler mit weiteren Komponenten, wie z.B. Schaltflächen, füllen kann. Der Entwickler muss zwei Interfaces anlegen. Ein Interface für die View sowie ein weiteres Interface (*CalculatorViewListener*), in der die Event-Methoden definiert werden. Dieses Interface wird im View-Interface platziert. Die konkrete View-Klasse erbt ein Interface (*CalculatorView*) und die benötigten Standard Vaadin-Listener, in diesem Fall den ClickListener. Das Interface ClickListener erfordert die Implementierung einer Methode, um auf einen Schaltflächenklick zu reagieren. Über den Konstruktor der Button-Klasse wird allen Schaltflächen im Beispiel die View als Listener zugeordnet. Sie ruft die Klick-Funktion aus dem ViewListener auf, der vom Presenter implementiert wird. Über die Schaltflächenbeschriftung, die übergeben wird, kann der Presenter die gewünschte Operation unterscheiden. [GM13]

Die vorgestellten Varianten ermöglichen die Nutzung von MVP mit Vaadin. An beiden lässt sich kritisieren, dass zunächst die View auf das originale Klick-Ereignis von Vaadin reagieren muss. Die Vor- und Nachteile der beiden Versionen sind zusammengefasst: Die Variante mit dem EventBus-System ist flexibel, verständlich und somit leicht umsetzbar. Zudem kann sie auch außerhalb des Vaadin Frameworks eingesetzt werden. Als Vorteil der Vaadin-Variante kann gesehen werden, dass eine zusätzliche Abhängigkeit verhindert wird. Allerdings ist die Umsetzung komplexer.

2.2 Darstellung von Aufgaben zu Dokumenten

Nach der technischen Realisierung der Trennung von Prozess- und UI-Steuerungslogik, behandelt dieses Unterkapitel die Darstellung einer Aufgabe mit sinnvollen Informationen aus einem Dokument im User Interface. Bereits die Erfassung nur weniger Daten zu einem Dokument kann sich in einem Anwendungssystem bereits als komplexe Herausforderung darstellen. Dies ist besonders dann der Fall, wenn sich die einzutragenden Angaben auf Daten beziehen, die bereits wesentlich eher im Verarbeitungsprozess erfasst wurden und somit nicht direkt ersichtlich sind. Zudem kann

das Problem auch dann bestehen, wenn ein Sachbearbeiter in einem mehrseitigen Formular nur an einer bestimmten Stelle die Daten erfassen muss. Um diesem Problem entgegen zu wirken, lassen sich die Daten des Formulars in zwei Kategorien einteilen. Zum einen gibt es Daten, die in dem Dokument bereits erfasst wurden. Sie dienen dem Sachbearbeiter als Wissensbasis und ggf. als Grundlage zur Erfassung weiterer Daten. Zum anderen müssen neue Informationen zu einem Dokument erfasst werden. Dabei kann es sich um Informationen, aber auch um Freigaben oder Ergebnisse einer Dokumentenprüfung handeln. Diese Erfassung von Informationen lässt sich als *Aufgabe* bezeichnen, die z.B. einem Sachbearbeiter zugewiesen ist. In einem Prozess nach BPMN 2.0 wird eine Aufgabe durch eine Benutzer-Aufgabe (User Task) definiert. Im Gegensatz zu einer manuellen Aufgabe kann sie durch eine Process Engine gesteuert und in dem Zuge einem Benutzer oder einer Benutzergruppe zugewiesen werden, indem sie in einer benutzerspezifischen Aufgabenliste (Task List) abgelegt wird. Zur Erledigung der Aufgabe muss eine entsprechende Interaktion, z.B. Eingabe oder Button-Klick, des Nutzers erfolgen. Erst dann wird der User Task beendet und der Prozess kann weiter verarbeitet werden. [FR12, S. 72]

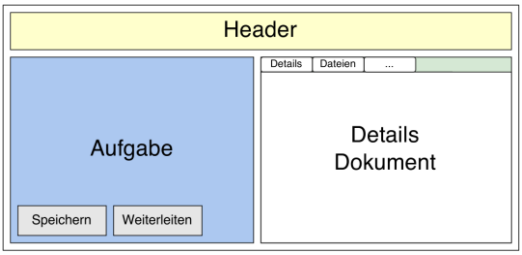


Abb. 3: Darstellung einer Aufgabe zu einem Dokument, angelehnt an [SN11, Abb.8-8]

Die User Tasks können aus der Process Engine abgefragt und auf einer Benutzeroberfläche (User Interface) dargestellt werden. Dies bietet dem Nutzer die Möglichkeit die Daten in einer Formularansicht zu erfassen und abschließend durch eine Schaltfläche zu bestätigen. Bei der Darstellung der Aufgabe sollte beachtet werden, dass benötigte Informationen zur Bearbeitung der Aufgabe entsprechend übersichtlich angezeigt werden. Es bietet sich an, die zu erfassenden und die benötigten Informationen nebeneinander zu platzieren (siehe Abb. 3:). Der linke Aufgabenbereich gestaltet sich entsprechend der jeweiligen Aufgabe, wohingegen der rechte Bereich das Dokument (Business Object) mit allen bereits erfassten Daten registerweise abbildet. Die Register sollten dabei in sich zusammenhängende Kategorien bilden und aufgabenspezifisch geöffnet werden. Besonders hervorzuheben ist, dass die Darstellung im rechten Bereich für alle Aufgaben identisch ist. Dies führt zu einer besseren Orientierung und Wiedererkennung. Um die gesamte Breite eines Bildschirms zur Visualisierung nutzen zu können, sollte zudem auf weitere Elemente in der horizontalen Ebene verzichtet werden. Dazu zählt auch ein vertikales Menü, das durch ein horizontales im oberen Bereich der Ansicht ersetzt werden sollte. Außerdem ist es hilfreich, die vertikale Trennung zwischen Aufgaben und Informationen dynamisch und variabel zu gestalten.

Dadurch lässt sie sich je nach Bedarf verschieben und die einzelnen Bereiche können gezielt vergrößert dargestellt werden.

Ein derartiges Vorgehen wird auch mit dem *Process/BO-Pattern* für BPM-Projekte beschrieben. Es sieht vor, dass zu jedem Prozess ein zentrales Business-Objekt (BO) gehört. Die aufgabenübergreifend wichtigsten Daten, wie z.B. ID, Bezeichnung und Ersteller, werden im Header abgebildet. [SN11, S. 192f] Eine Erweiterung des Patterns stellt das sogenannte *Process/BO-Portlet mit Task Data* dar. Sie definiert einen gesonderten Bereich für die Erfassung von Benutzereingaben in Form einer Art Formular. Daneben werden in einem weiteren Bereich die für die Erfassung relevanten vorhandenen Informationen angezeigt. [SN11, S. 280]

2.3 Mapping von User Task zu User Interface mit Vaadin und Camunda BPM

Nach dem grafischen Entwurf der Darstellung, muss dieser auch technisch umgesetzt werden. Es muss ein Mapping zwischen UI-Maske und User Task im Prozess realisiert werden. Zudem sollen die Masken, wie in Kapitel 2.2 beschrieben, als weitestgehend unabhängige und somit austauschbare Komponenten geladen werden. Der BPMN 2.0 Standard enthält keine Definitionen für das User Interface, daher unterscheidet sich das Vorgehen je nach Wahl der Process Engine. Ein Pattern zum Erstellen bzw. Laden von Objekten ist das Factory Design Pattern. Es bietet Methoden zum Erzeugen von Objekten von bestimmten Typen. Über Parameter entscheidet der Algorithmus der Factory, welches Objekt erstellt und welches zurückgeliefert wird.

In diesem Abschnitt werden der Einfachheit halber nur die Begriffe und Techniken von Camunda genannt und verwendet. Für das Mapping zwischen UI-Maske und User Task im Prozess existiert in Camunda BPM eine Variable namens FormKey. Hier kann bspw. der Pfad zu einer View-Komponente oder JSP-Seite hinterlegt werden. [CM13]

Die Darstellung der Aufgaben und des Dokuments erfolgt auf einer Webseite. Es gibt aufgrund der Verwendung des MVP-Patterns eine View und einen Presenter für die Bearbeitungsseite. Die User Tasks können über eine Vaadin CostumComponent dargestellt werden. In dieser Klasse können Standardkomponenten beliebig platziert und angesteuert werden. [GM14, Kap 5.29] Sie sind die Grundlage für die Aufgabenkomponenten. Der Presenter prüft beim Aufruf der Bearbeitungsseite, ob über die URL eine Identifikationsnummer eines Dokumentes übergeben wurde. Wenn dies der Fall ist, prüft er über das Model, ob es das Dokument gibt und ob dazu ein User Task vorliegt. Dann muss der Presenter dafür sorgen, dass die passende Aufgabenkomponente angezeigt wird. Die Instanziierung der passenden Aufgabenkomponente erfolgt nach dem Factory-Design-Pattern. Das trägt zu einem übersichtlichen Quellcode im Presenter bei und eröffnet die Möglichkeit, die Art der Instanziierung flexibel zu ändern, ohne den Presenter bearbeiten zu müssen. [CJ10]

Die Factory ist *static* deklariert und bietet eine Funktion zur Erzeugung der Aufgabenkomponenten. Als Übergabeparameter erwartet die Funktion den FormKey des

User Tasks aus der Prozessinstanz. Alle Aufgabenkomponenten erben von einer abstrakten *AbstractTaskComponent*-Klasse. Alle Klassen, die von der abstrakten Klasse erben, haben Zugriff auf den EventBus, der bei der Instanziierung über den Konstruktor gesetzt wird. Weiter sind zwei abstrakte Methoden definiert. Eine Methode zum Einlesen eines Datenobjektes, dessen Informationen in der UI-Komponente angezeigt werden soll, und eine Methode, die die Eingaben des Benutzers in das Datenobjekt zurückschreibt. Der Presenter braucht keine Kenntnis über den konkreten Typ der Aufgabekomponente, weil er die Komponenten ausschließlich mit den abstrakten Methoden der Oberklasse steuern kann. Die Schaltfläche zum Abschließen einer Aufgabe ist in der Aufgabenkomponente platziert. Somit empfängt sie das Klick-Ereignis und leitet es an den EventBus weiter. Die Subscriber-Methoden sind im Presenter platziert (siehe 2.1). Die Factory erzeugt die passende Instanz via Reflections. Der Einsatz von Reflections ist risikobehaftet. Es können Exceptions auftreten, die bei einem „normalen“ Instanziieren einer Klasse nicht auftreten. Es sollten Maßnahmen getroffen werden, um die Risikoquellen gering zu halten: [OR14] Der FormKey wird nicht dazu genutzt, um den Pfad der konkreten Klasse zu hinterlegen, sondern fungiert als Key für die Map in der *TaskComponentFactory* (siehe Abb. 4). Nur wenn zu dem übergebenen Key eine Klasse hinterlegt ist, startet die Instanziierung. Ansonsten wird dem Benutzer eine Fehlermeldung angezeigt und die Entwickler über Logfile-Einträge informiert. So kann sichergestellt werden, dass bei einer Umbenennung der Komponente, beispielsweise im Rahmen eines Refactorings, das Prozessmodell und laufende Prozessinstanzen weiterhin funktionieren, ohne Exceptions zu erzeugen.

Eine weitere mögliche Fehlerquelle ist die Hinterlegung von unpassenden Klassen in der Map. Damit nur Klassen vom Typ *AbstractTaskComponent* hinterlegt werden können, ist der Typ-Parameter der Map mit dem Schlüsselwort *extends* definiert.

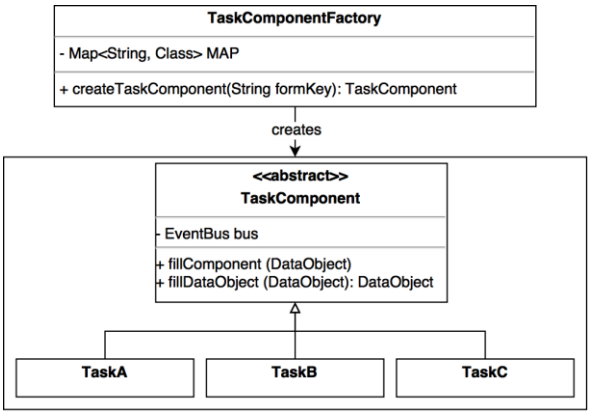


Abb. 4: UML-Klassendiagramm mit Factory und Vererbungsstruktur der Aufgaben

Ein neuer User Task erfordert nach diesem Ansatz lediglich die Entwicklung einer Aufgabenkomponente sowie das Hinzufügen eines Key-Value Paars in die Map. Diese

Lösung erlaubt die Umbenennung von Klassen, ohne die Prozesse anpassen zu müssen. Zudem kann der Compiler erkennen, ob die in der Map eingetragenen Klassen existieren und Typsicherheit gewährleisten. Die typischen Reflections-Probleme, die erst zur Laufzeit auftreten, werden so weitestgehend vermieden.

Eine weitere Möglichkeit wäre es, die Map aus einer Konfigurationsdatei zu erzeugen. Das hätte den Vorteil, dass die Komponenten im Betrieb auch für laufende Prozessinstanzen ausgetauscht werden können, z.B. wenn eine überarbeitete Version vorliegt, die einfacher für den Benutzer zu bedienen ist. Diese Variante bietet keine Typsicherheit und bei einem Refactoring muss die Konfigurationsdatei ggf. überarbeitet werden, um die Lauffähigkeit der Prozesse sicherzustellen. Ein Wenn-Dann-Konstrukt kommt nicht in Frage. Es ist starr und wird bei einer größeren Anzahl von User Tasks unübersichtlich. Eine Änderung der Konstruktor-Parameter beträfe mehrere Programm-Zeilen, während bei der Reflections-Variante nur eine Zeile verändert werden muss.

Auf der Webseite von Vaadin wird gezeigt, wie das Framework mit der Process Engine Activiti, von der Camunda BPM ursprünglich abstammt [NA13], eingesetzt werden kann. Das Vaadin-Beispiel setzt auch auf Reflections in Kombination mit einer typsicher gestalteten Map. Die Klasse *UserTaskFormContainer* aus dem Beispiel agiert auch hier als Fabrik und instanziiert die Objekte. Der Unterschied zur oben dargestellten Variante ist, dass die Klassen zur Laufzeit über eine Methode zur Map hinzugefügt werden können. Außerdem werden im Vaadin-Beispiel nicht Komponenten in eine Seite geladen, sondern es erfolgt eine Weiterleitung zur entsprechenden Seite. [MP11]

3 Process Patterns

Im Folgenden werden prozessorientierte Patterns im Zusammenhang mit der Verarbeitung von Dokumenten im BPMN Kontext dargestellt. Im Einzelnen wird näher auf die Erfassung von Dokumentenstatus sowie die gleichzeitige Verarbeitung mehrerer Dokumente in einem BPMN-Geschäftsprozess eingegangen.

3.1 Implementierung eines Statuskonzeptes in BPMN

Ein Dokument kann innerhalb eines Genehmigungsprozesses unter Umständen viele unterschiedliche Abteilungen oder gar Unternehmen durchlaufen. Die Definition eines Statuskonzeptes ermöglicht einen besseren Überblick und den gesamten Workflow. Dadurch lässt sich zu jeder Zeit der Zustand einer Prozessinstanz und damit der eines Dokumentes feststellen. Außerdem ist ersichtlich, welche Person der vorangegangene bzw. nachfolgende Sachbearbeiter ist.

Bei dem Einsatz von BPMN wird der Schwerpunkt auf den Prozessablauf (Sequenzfluss) gelegt. Andere Aspekte, wie z.B. Daten und Datenfluss, werden zweitrangig behandelt. [FR12, S. 103] Dennoch sind für die Abbildung des Status eines

Dokumentes in einer BPMN-Anwendung unterschiedliche Umsetzungsmöglichkeiten denkbar. So bietet die Modellierung von Datenobjekten (Data Object) innerhalb des Diagramms eine Art der Darstellung von Informationen. Ein Datenobjekt kann an der entsprechenden Stelle im Sequenzfluss platziert und neben der Bezeichnung auch ein Status in eckigen Klammern definiert werden (siehe Abb. 5:). Es legt dar, welche Daten zu einem bestimmten Zeitpunkt in Prozess vorhanden sein müssen und welchen Status sie besitzen sollten. Dementsprechend verhält es sich mit Dokumenten, die durch eine Aufgabe (Task) erzeugt oder manipuliert werden. Dies unterstützt den Leser eines BPMN-Diagramms durch die Anreicherung von Details und führt zu einem besseren Verständnis des Prozesses.

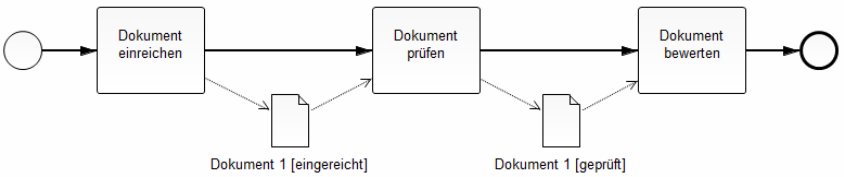


Abb. 5: Statuserfassung mit Datenobjekten in BPMN 2.0

Trotz der unterstützenden Funktionen bietet dieser Ansatz einen entscheidenden Nachteil: Es ist keine automatisierte Bearbeitung des Datenobjektes sowie dessen Status im Rahmen von ausführbaren BPMN-Modellen möglich. Wenn sich beispielsweise durch einen Task der Status eines Dokumentes von *eingereicht* in *geprüft* ändert, ist dieses im Diagramm zwar abbildbar, der Status eines Dokumentes kann jedoch nicht explizit abgefragt werden. Dafür wäre es notwendig den Status separat, z.B. in einer Datenbank, zu erfassen.

Eine alternative Möglichkeit bietet der Einsatz von Zwischenereignissen (Intermediate Events). Sie können durch den Prozess ausgelöst werden und einen Status beschreiben, der durch das Durchlaufen des Sequenzflusses erreicht werden kann. [FR12, S. 26]

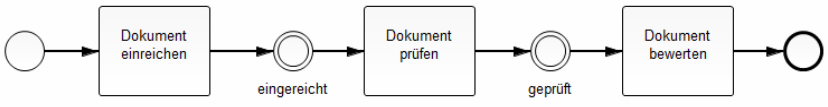


Abb. 6: Statuserfassung mit Intermediate Events in BPMN 2.0

In dem beispielhaften Prozess, der in Abb. 6: dargestellt ist, wird zunächst ein Dokument eingereicht. Unmittelbar im Anschluss daran folgt ein Intermediate Event, das den Status des Dokumentes als *eingereicht* kennzeichnet. Ebenso verhält es sich mit der Aktivität *Dokument prüfen* und dem darauf folgenden Status *geprüft*.

Durch die Zwischenereignisse lässt sich demnach, vergleichbar mit den Datenobjekten im Diagramm, der Status des Prozesses bzw. des darin verarbeiteten Dokumentes erkennen. Jedoch bieten die Events zusätzliche Funktionalitäten für die Verarbeitung in

ausführbaren BPMN-Anwendungen mit Camunda. So ist es möglich zu einem Ereignis einen Listener zu definieren und einen Ausdruck (Expression) zu hinterlegen, der bspw. das Aufrufen einer Methode bewirkt. Der Methode wird der entsprechende Status mit übergeben, sodass dieser zum Dokument gespeichert werden kann.

Innerhalb der Methode *veraendereStatus* wird zunächst das Dokument über eine ID ermittelt, die sich über die *execution*-Variable aus dem Prozess auslesen lässt. Zu dem betreffenden Dokument wird anschließend der Status in einer Datenbank erfasst. Der entscheidende Unterschied zur Umsetzung mit Datenobjekten ist, dass der Methodenaufruf mit dem entsprechenden Status direkt aus dem Modell erfolgt. Die Methode ist einmalig zu entwickeln, danach kann sie in jedem Intermediate Event verwendet werden und es sind keine weiteren Programmierungen außerhalb des Diagramms mehr notwendig. Dies hat den Vorteil, dass Änderungen und Erweiterungen im Prozess vereinfacht werden. Zudem erfolgt für Beteiligte des operativen Prozesses eine erhöhte Transparenz. So lässt sich der Zustand der Prozessinstanz und damit der Status eines Dokumentes zu jeder Zeit, z.B. über ein Verwaltungssystem, prüfen.

3.2 Mengenverarbeitungsprozess

Bei dem Einsatz einer BPMN-Anwendung werden Prozessinstanzen erzeugt. Dies geschieht, wenn in der Realität ein neuer Prozess gestartet wird [FR12, S. 23]. Somit kann man unter einer Instanz einen Prozess verstehen, den z.B. ein Dokument zur Bearbeitung durchläuft. Dieses wird durch unterschiedliche Sachbearbeiter solange editiert, bis alle erforderlichen Daten vorhanden sind. Dabei kann es vorkommen, dass zu einem bestimmten Ereignis mehrere Dokumente mit dem gleichen Status am selben Wert manipuliert werden. Dieser Fall tritt bspw. ein, wenn im Rahmen einer Sitzung Entscheidungen über die Prioritäten einer Vielzahl von Anträgen getroffen werden. Das Ziel ist die Datenerfassung über mehrere Anträge auf nur einer Ansichtsmaske, z.B. in einer Tabelle. Dieser Abschnitt beschäftigt sich mit der Frage, wie sich mehrere unabhängige Prozessinstanzen gebündelt innerhalb eines Tasks weiterbearbeiten lassen.

Die Verarbeitung des Prozesses wird in dem o.g. Beispiel durch das Einreichen eines Dokumentes initialisiert. Im BPMN-Prozess durchläuft die entsprechende Prozessinstanz verschiedene Tasks und Gateways, bis das Ereignis eintritt, in dem über eine Menge von Dokumenten gebündelt entschieden wird, z.B. im Anschluss an eine Bewertung des Dokumentes. An dieser Stelle lässt sich im BPMN-Prozess ein Intermediate Event platzieren, wodurch die eintreffenden Instanzen so lange warten, bis die Erfassung der Daten in den Dokumenten erfolgt ist. Als Ereignistyp wird ein Nachrichtenereignis (Message Event) eingesetzt, da sich so die Kommunikation zwischen der Einzel- und Mengenverarbeitung abbilden lässt (siehe Abb. 7).

Die Mengenverarbeitung erfolgt in einem gesonderten BPMN-Pool, der ausschließlich für diese Tätigkeit verwendet wird. Die Initialisierung einer Instanz in diesem Pool kann durch eine Nutzerinteraktion über das User Interface erfolgen, z.B. durch das Anklicken einer Schaltfläche. Auf der Benutzeroberfläche wird dem User eine Übersicht aller

Dokumente dargestellt, in der er eine bestimmte Eigenschaft zu allen Dokumenten erfassen kann. Durch das Anklicken der besagten Schaltfläche, z.B. Speichern, erfolgen mehrere Teilschritte, die sich zu zwei Schritten zusammenfassen lassen.

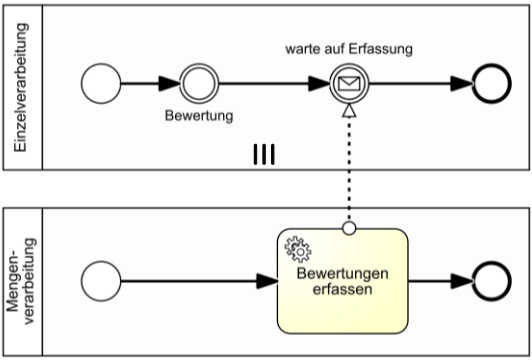


Abb. 7: Mengenverarbeitung mit Hilfe von zwei Pools in BPMN 2.0

Schritt 1: Datenerfassung in der Datenbank: Für alle dargestellten Dokumente werden die angegebenen Daten in einer Datenbank erfasst. Im o.g. Beispiel wären dies die Prioritäten zu den jeweiligen Dokumenten.
Schritt 2: Aufgabe zur Datenerfassung abschließen: Zunächst wird ein neuer Prozess in der Mengenverarbeitung instanziiert. Daraufhin wird die Aktivität *Bewertungen erfassen* in Form eines Service Tasks bearbeitet. Die Ausführung des Service Tasks erfolgt automatisiert. Sie versendet eine Nachricht, auf die das Message Event der jeweiligen Prozessinstanzen im Einzelverarbeitungsprozess (Gesamtprozess) wartet (s.o.).

Ein Problem dabei stellt die Transaktionssicherheit dar. Ein konsistenter und widerspruchsfreier Zustand ist für die Geschäfts- und die Prozessdaten nicht zu jeder Zeit gewährleistet. Sollte nach Schritt 1 z.B. ein Fehler auftreten und die Verarbeitung abbrechen, sind die Daten in der Datenbank erfasst. Der Prozess bildet dies jedoch nicht ab, da die Aufgabe nicht abgeschlossen wurde und die Prozessinstanzen weiterhin beim Message Event auf eine Nachricht warten.

4 Fazit

Probleme, die bei der Entwicklung einer prozessgetriebenen Anwendung auftauchen, können häufig durch Standardpatterns gelöst werden. Etablierte Patterns haben in der Praxis bewiesen, dass sie ein Problem gut und sinnvoll lösen können. Die Verwendung führt zu einer übersichtlicheren Architektur. In der Architekturbeschreibung kann auf Patterns verwiesen werden und neuen Entwicklern fällt die Einarbeitung leichter.

Die in diesem Paper vorgestellten Patterns in Kapitel 2.1 (Anwendung des Model View Presenter Patterns mit Vaadin) und 2.3 (Mapping von User Task zu User Interface mit

Vaadin und Camunda BPM) sind auf Vaadin und Camunda zugeschnitten.

Die Patterns zur Gestaltung der Benutzeroberfläche, die Implementierung eines Statuskonzeptes in BPMN und der Mengenverarbeitungsprozess können, wie bei Patterns üblich, mit geringen Anpassungen in jede Prozess-Anwendung übernommen werden. M. Fowler schrieb, dass Patterns nicht "blind" übernommen werden sollen, sondern "im Ofen ihres Projektes nachgaren" müssen. Die Verwendung von Patterns ohne Anpassungen an den Kontext führe zu schlechten Ergebnissen. [Fo09, S. 24]

Literaturverzeichnis

[CM13] Camunda Services GmbH: Tutorials & How Tos - User Interface, http://docs.camunda.org/latest/real-life/how-to/; Stand: 26.02.2015.

[CJ10] Cooper, J.: Java Design Patterns: A Tutorial. Addison Wesley, Reading Massachusetts, 2010.

[Fo06] Fowler, M.: http://www.martinfowler.com/eaaDev/uiArchs.html#Model-view-presentermvp; Stand: 26.02.2015.

[Fo09] Fowler M: Patterns für Enterprise-Application-Architekturen. MITP-Verlag, Heidelberg, 2003.

[FR12] Freund, J.; Rücker, B.: Praxishandbuch BPMN 2.0. Hanser, München, 2012.

[GM13] Grönroos, M.: Model-View-Presenter Pattern with Vaadin. https://Vaadin.com/web/magi/home/-/blogs/2342730; Stand: 26.02.2015.

[GM14] Grönroos, M.: Book of Vaadin - Vaadin 7 Edition. 4. aktualisierte Auflage. Vaadin, 2014.

[MP11] Holmström, P.: Building Vaadin Applications on top of Activiti, https://vaadin.com/wiki?p_p_id=36&p_r_p_185834411_title=Building+Vaadin+Applications+on+top+of+Activiti; Stand: 25.02.2015.

[NA13] Neumann, A.: Camunda stellt BPM-Plattform unter Open-Source-Lizenz, http://www.heise.de/developer/meldung/camunda-stellt-BPM-Plattform-unter-Open-Source-Lizenz-1824467.html; Stand: 26.02.2015.

[OR14] Oracle: Trail: The Java Tutorials - The Reflection API, http://docs.oracle.com/javase/tutorial/reflect/; Stand 25.02.2015.

[SN11] Slama, D.; Nelius, R.: Enterprise BPM. dpunkt.verlag, Heidelberg, 2011.

[WL12] Wasseman, L.: Event Bus Explained, https://code.google.com/p/guava-libraries/wiki/EventBusExplained; Stand: 25.02.2015.

[VA15] Vaadin Ltd.: Introduction, https://Vaadin.com/introduction; Stand: 26.02.2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Game Event Lenses: Focus+Context Visualizations of Computer Game Data

Lars Schütz[1]

Abstract: This paper introduces Game Event Lenses. They enable explorative and interactive focus+context visualizations of computer game events that may form huge and complex datasets including various attributes from the close game environment. The Data State Reference Model (DSRM) is the foundation of the presented approach. It is based on different data states and operators. This work features a concept that is designed to completely simulate the DSRM on a graphics card by using its memory and related shader programs. Their new general purpose computing and traditional rendering capabilities allow user interactions and frequently generated images in real time.

Keywords: Information visualization, graphic systems and interfaces, computer game data.

1 Introduction

Computer games are versatile, technically complex and widely spread. They can be differently perceived. Developers might create new virtual worlds and adventures with dedication and passion. They and additional publishers try to make a profit with their games while players typically consume them to have fun or to simply pass time. There are more concrete motivational points when it comes to games.

Developers and publishers try to analyze the behavior of players in solo and group activities [EDC13]. By doing this, they possibly gain insight into specific trends and preferences, e. g., favorite in-game weapons. Furthermore, they might detect if the players really play their game how they meant it to be played. Such obtained information can result in further adjustments to the game and its mechanics that even might get into the game after its initial release. Individual short term offers for in-game items that can be bought with real money or layout changes of specific in-game world areas can be seen as only two related of many more imaginable examples.

Other motivations from a developer's point of view refer to the general game design. Weak points regarding the overall game balance can become obvious by analyzing each player's progress and performance in a game. For example, several players have to fight enemies and no one is able to beat them because their attack skill is numerically too weak. Those kind of shortcomings can possibly be seen when collected game data is examined. Such an analysis can lead to game balance adjustments, e. g., lowering the overall difficulty level.

[1] Anhalt University of Applied Sciences, Department of Computer Science and Languages, 06366 Köthen, Germany / Otto von Guericke University, Faculty of Computer Science, 39106 Magdeburg, Germany, l.schuetz@inf.hs-anhalt.de

Players view games from another perspective. They may take part in electronic sports tournaments where the best strategies and tactics typically lead to a tournament win. In this respect, they possibly have a high interest in the analysis of their performances. But also ordinary players compare themselves by browsing player profiles and statistical overviews about achievements and owned items in various games. A game can almost always be seen as a competitive environment.

Generally speaking, the analysis of computer game data involves different directions and perspectives. The afore-mentioned approaches only give a basic overview. A separate discipline called Game Analytics evolved during the last years that deals with such questions [EDC13]. It combines several other subjects like Data Mining, Machine Learning, Statistics and Visualization to gain insight into game related data.

1.1 Challenges

The extraction and communication of usable information from huge datasets is a central issue, because it is difficult to deduce structural information and interrelationships from raw data representations. It can be a very time-consuming and complex task. The visualization of the raw data may help in this respect. Additionally, supporting tools are needed that process data in a timely manner.

A consequent problem is the possible information overload when many data elements are visualized at the same time. That complicates the recognition and interpretation of facts. Specific visualization techniques may help to address that challenge.

All those issues affect visualization in general and they are applicable to the visual analysis of computer game data in particular. This paper examines possibilities to overcome those problems and therefore provides a useful contribution to the specific area of Game Analytics as well as to the discipline called Information Visualization [Sp01].

1.2 Outline

The visualization of raw data makes the illustration of information possible and more accessible. It is seen as one chance to address the previously mentioned challenges. Section 2 clarifies what visualization actually means and how initial data can be processed by introducing the visualization transformation and the Data State Reference Model (DSRM) [CR98].

Section 3 refers to (computer) game events in the context of the DSRM. Game events are defined, their components are identified and it is outlined how they can be processed.

Visualizations of game events in particular and of data in general still might lead to cluttered information as previously mentioned. This hinders the discovery of new insights into the data. Regarding this, lenses are presented in Section 4. They enable a user to focus on specific parts of a visualization.

To process and to visualize the game events with respect to the integration of lenses a performant graphics system is needed. Section 5 shows the conceptual idea of such a system and delivers introductory examples based on a collected sample dataset of game events.

Finally, Section 6 concludes this paper and proposes future work.

2 Visualization and the Data State Reference Model

A visualization is typically understood as an external graphical representation of concepts or data [Wa04]. In this respect, it stands for a final result of some kind of process. In contrast to that, another point of view refers to a visualization as a human process itself that leads to a mental model of data [Sp01]. This approach emphasizes the human brain with its perceptive and cognitive abilities. We can combine both views and see a visualization as an interface between them [Wa04]. Humans with their comprehensive faculty and decision making abilities are on the one side. Computers that process huge amounts of data, e. g., images from the internet, to further abstractions by using specific algorithms are on the other side. A visualization is then the foundation for the knowledge acquisition. It carries information that is based on the given data that should be easier to understand. This whole idea is illustrated in Fig. 1.

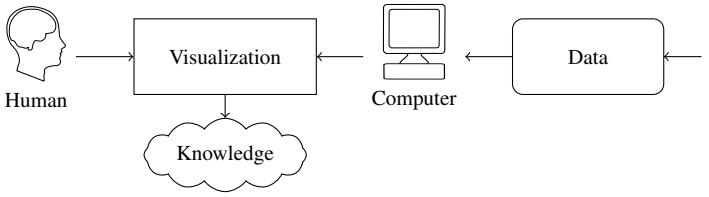


Fig. 1: Visualization acts as an interface between the human and the computer-processed data and enables the discovery of knowledge

It now needs to be clarified how visual representations or image data D_I are generated starting with available raw data D_R before we are actually able to derive knowledge from them. The involved process is the so called visualization transformation that is basically the mapping $f : D_R \times P_1 \times P_2 \times \cdots \times P_n \rightarrow D_I$ with $P_i \mid 1 \leq i \leq n, n \in \mathbb{N}$ being the visualization parameters and $p \in P_i$ being the visualization parameter values [JMG02]. This parameterized form gives us more control and flexibility because each set P_i can be applied to different steps in the visualization transformation depending on how it is actually defined.

The DSRM describes how raw data can be transformed into image data. It is based on different data states and operators. The following four states, also called stages, need to be distinguished:

- *Raw data* D_R are the input data for the visualization.
- *Analytical abstractions* D_A are calculated properties or meta data.

- *Visual abstractions* D_V are abstracted data in portrayable form that carry visual properties, e. g., color hue values.
- *Image data* D_I are the final visual representations, e. g., pixel data.

The DSRM works like a pipeline. It consists of two types of operators that propagate the data through the model:

- *State operators* leave the structure of the data untouched because the output is on the same stage as the input. There are operators O_R on D_R, operators O_A on D_A, operators O_V on D_V and operators O_I on D_I.
- *Transformation operators* transform data from one stage to their succeeding stage. They can be distinguished into the following operators:
  - *Raw data transformations* T_R abstract raw data and create new structures.
  - *Visualization transformations* T_V map analytical abstractions to visual abstractions.
  - *Visual mapping transformations* T_A render visual abstractions to obtain image data.

The first two states are part of the data domain while the remaining states are part of the view domain. The same principle applies to the operators. This kind of classification enables a distinction between data-oriented or semantic operators and view-oriented or graphical operators [TFS08]. The schematic view of the DSRM is depicted in Fig. 2.

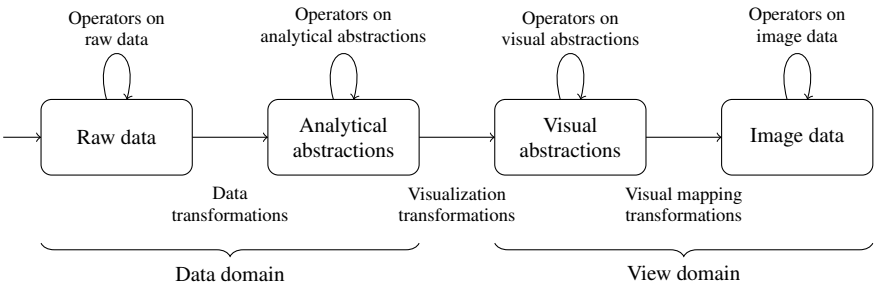


Fig. 2: The DSRM transforms raw data via analytical abstractions and visual abstractions into image data by using operators

3 Game Events in the Context of the Data State Reference Model

The following examinations apply the DSRM to game events. Aspects that relate to the data domain and the view domain are considered. But first of all the most important terms shall be clarified.

A *(computer) game* is an interactive experience that offers increasingly challenging sequences of patterns to a player that can learn and potentially master those [Ko04]. We can conclude that the gaming experience can turn out differently for various persons while the patterns may be the same. That relates to the individual learning aptitude of each player and to the fact that they interact with or within the game in different ways, e. g., in terms of chronological order.

The term *event* is versatile and actually defined differently in various disciplines. For example, events are essential when it comes to interrupt processing and the process life cycle in operating systems [Ma13]. In that context, events trigger temporary and final interrupts of process executions to handle new or already existing processes. Those kind of events let a system react to them. In discrete event-oriented simulations events are instant occurrences that alter a system's state [To11]. Such a state comprises a collection of the system's properties at a specific time. We can find different general concepts for an event that are sometimes used interchangeably. On the one hand, there are event types or event classes that describe different kind of events and determine a unified structure for similar events while, on the other hand, there are event instances or event objects that are a concrete realization or instantiation of an event type [BD10]. In this paper, an event stands for an event instance unless stated otherwise.

A *(computer) game event* can now be defined as a detectable and immediate event that contains parts of the complete game state description and is triggered by game mechanics or internal requests. This definition focuses on the following aspects:

- *Detection:* A game event can be detected by the game itself. It can further be detected by either nobody or at least one player.
- *Duration:* A game event occurs instantly and does not span a specific period of time.
- *State description:* A game event represents properties of game related elements. It is an excerpt from the game's current complete internal state description.
- *Game mechanics or requests:* On the one hand, at least one action of a player triggers an event. Respectively, an action is caused by interacting with the game, e. g., pressing a button to move a virtual avatar. That refers to game mechanics in the sense of gameplay. On the other hand, the game's internal processes trigger events, too. Those are either game mechanics with no players involved, e. g., jumping of a non-player character, or procedures that are relevant to the system itself, e. g., getting the number of currently logged in players in an online game.

3.1 Game Events in the Data Domain

Game events consist of different components that represent the game's partial state description. They form the input of the DSRM. We propose the following classification:

- The *event type* describes the event's class or the type of the corresponding event instance.

- The *timestamp* informs about the point in time the event happened, e. g., by stating date and time information.
- The *spatial data* describe several spatial information, e. g., the position or place of the happening in in-game world coordinates or orientations of game entities.
- The *sender* holds information about the event's trigger.
- The *receiver* may store associated data if the event affects a game related entity. The receiver can also be the sender at the same time.
- The *event type data* are specific information based on the event type.

These components can result in many game events with different concrete attributes. They are flexible and suit different fields of application, e. g., a gameplay or business context. Tab. 1 shows two examples for a game event to further substantiate the appropriateness of the mentioned components.

Component	Gameplay game event	Business game event
Event type	Attack	Transaction
Timestamp	2015-03-07T23:01:14Z	2015-01-18T11:11:11Z
Spatial data	Area: 23, x: 234, y: 102, z: 20	Auction house: 15
Sender	Player: 3, level: 80, class: 4	Account: 66
Receiver	Player: 5, health points: 34; Player: 14, health points: 56	Account: 66
Event type data	Damage to player 5: 11; Damage to player 14: 5	Item: 54, quantity: 1, price: $7.99

Tab. 1: Two sample game events from different contexts

It is obvious that just one event can hold plenty of data depending on the degree of detail. There is no general rule. Each game may handle events differently. Some components may be optional. The shown arrangement and hierarchical structure of the components are not mandatory, e. g., the spatial data could be part of the sender's data instead. Each component further divides itself into several attributes, e. g., the receiver's health points. The proposed components are a working approach to describe the raw data but each visualization system should be aware of the complexity and possible different structures.

It is now clear how the raw data looks like. Several different operators can be used to further process the game events depending on the goal or task. Data cleansing, data reduction or data transformation are typical activities with a lot of different specific operators, e. g., automatic or manual selection of events (O_R or O_A), dealing with missing attribute values (O_R or T_R), detection and handling of outliers (O_R), normalization and standardization (O_R or O_A), attribute selection by principal component analysis (T_R), aggregation or discretization (O_R, T_R or O_A) [HKP12]. The operators may change the structure of the data so that analytical abstractions result from their application, or the data is simply passed through to the visualization transformation operators.

3.2 Game Events in the View Domain

The visualization transformation operators choose at least one of the visual variables from Bertin [Be11] to encode the analytical abstractions. Those variables are called *position*, *size*, *shape*, *brightness*, *color hue*, *orientation* and *texture*. It is possible to pick a visual variable for an attribute of a game event that should be transformed by observing its level of measurement. There are concepts that classify the degree of appropriateness for a visual variable, e. g., see MacEachren [Ma04]. Each game event or its further abstraction is represented by a geometric primitive with different levels of detail, e. g., a game event might have a position and a specific shape to encode its spatial data and type components.

The visual mapping transformation includes typical rendering operators, e. g., transformation and shading of geometric primitives, perspective or parallel projection and rasterization. The visual abstractions of the game events in form of geometric primitives are transformed into image data. Game events and possible further abstractions of them are now visualized. Operators for exemplary graphical manipulations on the image data are the translation, the rotation or the scaling of the viewable area.

4 Visualization Lenses for Game Events

The interpretation of visualizations can still be difficult because of some kind of information overload. Visualized game events may overlap each other. Too many game events may be illustrated at the same time. Therefore it is sometimes useful to concentrate on specific areas of a graphical representation. That area can include more detailed information while the surrounding remains simple and does not distract. That leads to focus+context visualizations by means of visualization lenses.

A visualization lens is an interactive and parameterizable visualization tool that spatially selects and modifies the underlying original representation of a visualization and helps to enrich, suppress or alter the corresponding content [To14]. The spatial selection represents the focus the user is interested in. The remaining part specifies the context that helps to orientate oneself. That enables a user to focus on different regions one after the other. In order to do this, the interaction with the lens is necessary that can be achieved by controlling its properties *shape*, *position*, *size* and *orientation* [To14].

A lens needs to be integrated into the visualization transformation. The focus region can be visualized by an additional DSRM pipeline that is attached to the main DSRM pipeline [To14]. A selection of the data in the focus region is needed. It is possible to select data from any stage of the DSRM, but a proper inverse projection of the selected data elements to their corresponding representations from previous states might be needed depending on the first operator in the lens pipeline. A management of unique identifiers for data elements or the concept of half spaces [TFS08] are resolutions for this problem. The lens determines the content of its focused region by realizing a composition $f = o_n \circ o_{n-1} \circ \cdots \circ o_1$ of $n \in \mathbb{N}_{\geq 1}$ operators o_i so that f can be interpreted as at least one state or transformation operator of the DSRM. The final output or the lens effect is joined with the main pipeline. Fig. 3 shows the described idea.

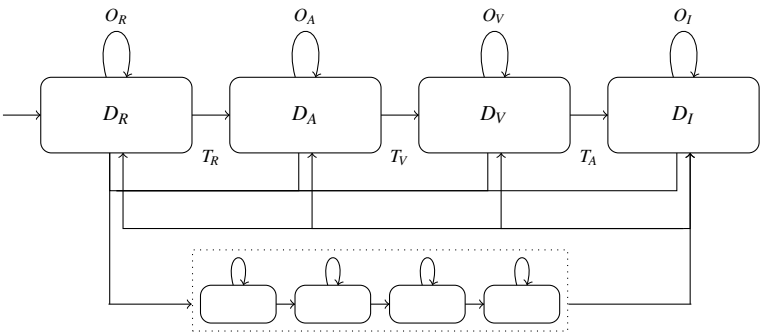


Fig. 3: A separate lens pipeline is attached to the main pipeline

Finally, the overall visualization of game events including the lens and its rendering is generated in three steps:

1. Create and render the base or context visualization by omitting the lens region.
2. Create and render the lens or focus region.
3. Optionally render the lens itself to show its properties, e. g., by drawing its boundary.

5 Graphics System and Prototypical Implementation

A software application is needed to explore game events. It is mainly responsible for the graphical representation and the processing of user interactions that alter the visualization parameter values. It should be able to process thousands of events in real time.

The proposed graphics system is based on the Open Graphics Library (OpenGL) [Sh13]. It is a cross-platform application programming interface and enables the control of the rendering process through the use of shader programs that operate on the graphics processing unit (GPU) and therefore profits from the GPU's parallel computing capabilities [Ow05]. OpenGL combines different shader types in a pipeline whereby each one has its distinctive functionality [Sh13]. These are in order the *vertex shader* (VS), the *tessellation control shader* (TCS), the *tessellation evaluation shader* (TES), the *geometry shader* (GS) and the *fragment shader* (FS).

In the presented concept the shader programs are used to simulate the DSRM's operators. Traditionally, vertices of geometric primitives are transformed into pixel data using at least a vertex shader and a fragment shader. This approach is similar to what the DSRM's graphical operators do. Consequently, those are covered and only the semantic operators need to be handled. Since August 2012 respectively OpenGL Version 4.3 the OpenGL pipeline is additionally equipped with *compute shaders* (CS) [Sh13]. They allow general purpose computations on the data that can be visualized afterwards. No extra interoperability efforts in terms of context switches are needed. A compute shader is used to simulate a semantic operator. Fig. 4 shows the proposed applicability of the shaders.

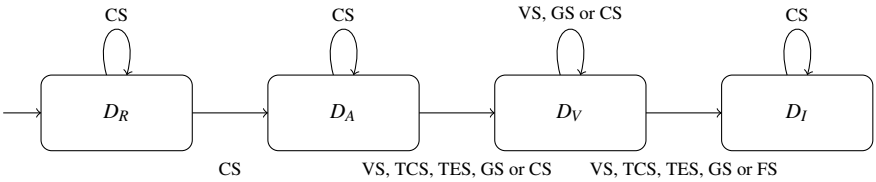
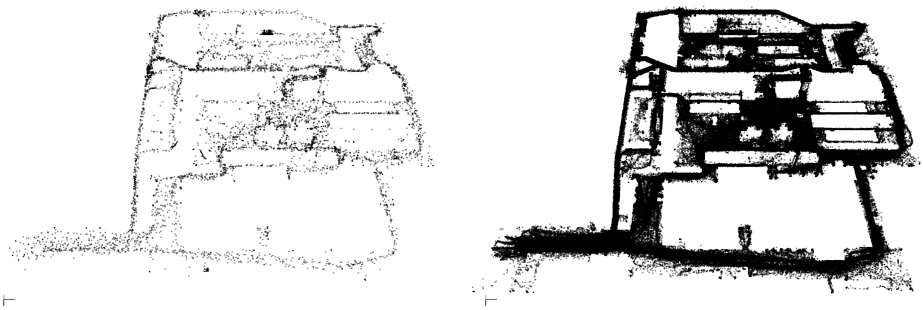


Fig. 4: Conceptual applicability of the shader types to simulate the DSRM's operators

In order to use GPU-based operators the data they work on need to be stored on the graphics card. All game events and further abstractions are coherently packed using *shader storage buffer objects* because shader programs can read from and write to them [Sh13]. That means that initially m game events with a maximum number of n attributes are stored in the sequence $(v_a^e) = (v_1^1, v_2^1, \ldots, v_n^1, v_1^2, v_2^2, \ldots, v_n^2, \ldots, v_1^m, v_2^m, \ldots, v_n^m)$ where v_a^e is the value of the attribute a that belongs to the specific event e. The visualization parameter values can be stored in *uniforms* or in *uniform buffer objects* [Sh13].

As a proof of concept two lens examples were developed and integrated into a software prototype that enables a user to interactively explore an example dataset, e. g., by moving the mouse to change the lens' position. The visualization design of the following examples is not addressed consciously and almost arbitrary, because only the feasibility of the shader-based approach is considered. That should definitely be examined and discussed separateley, e. g., in terms of choosing color hues or picking different shapes. The dataset originates from the computer game *Counter-Strike: Global Offensive*. The game events were collected on a private dedicated server applying a custom-built server plugin. 782000 events with 33 attributes for each event of overall 827 game rounds between two teams of computer controlled bots were logged. That corresponds to the play time of about 20.5 h. As a basic example Fig. 5 shows the spatial positions of each game event's sender component. This is the base visualization with no omitted events because there is no active lens.



(a) 13 game rounds (b) 827 game rounds

Fig. 5: Processed game events reveal the spatial position of their sender component

The first lens example allows to explore the category of the used weapon when a shot was fired by a player in relation to the event's position. The goal is to reveal preferred weapon

types. The lens starts operating on D_R and uses a CS $o_1 : D_R \rightarrow D_A$ (selects and passes the related events through), a VS $o_2 : D_A \rightarrow D_V$ (sets the position, and sets the color hue based on the weapon's category), a GS $o_3 : D_V \rightarrow D_V$ (changes the shape from point to triangle or square to encode the bot's team attribute) and a FS $o_4 : D_V \rightarrow D_I$ (passes the color through). Fig. 6 shows this example for one team by using a cuboid-shaped lens.

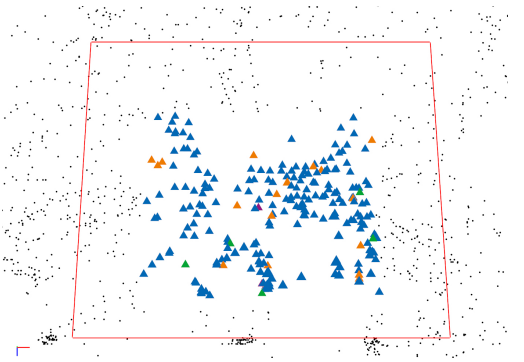


Fig. 6: Used weapon categories (encoded by different color hues) of one team when a shot was fired

The second lens example shows the positions of virtual player deaths and kills and the connections between them. The goal is to show dangerous and advantageous positions, and the related view directions at the same time. The lens starts operating on D_R and uses a CS $o_1 : D_R \rightarrow D_A$ (selects and passes the related events through), a VS $o_2 : D_A \rightarrow D_V$ (sets the position, and sets the color hue based on the team attribute), a GS $o_3 : D_V \rightarrow D_V$ (changes the shape from point to triangle (kill) or cross (death)) and a FS $o_4 : D_V \rightarrow D_I$ (passes the color through). The lens also uses a VS $o_5 : D_A \rightarrow D_V$, a GS $o_6 : D_V \rightarrow D_V$ (changes the shape from point to line) and a FS $o_7 : D_V \rightarrow D_I$ in a second pass to draw the connections. Fig. 7 shows this example with a sphere-shaped lens.

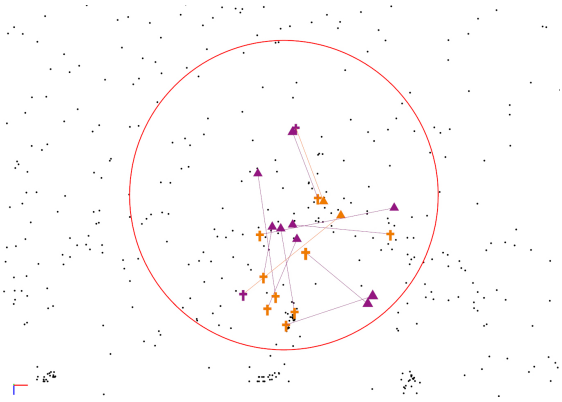


Fig. 7: Death and kill positions of both teams (encoded by shapes and color hues) and their relations

From a user's point of view performance is an important criteria in interactive applications. Loading times or processing times in general need to be minimized if a user interacts with

the system and when the visualization almost constantly changes, e. g., when adjusting the lens' position. An interactive experience is still possible with a maximum latency of about 66.667 ms that correspond to about 15 frames per second whereby a higher frame rate respectively a lower latency is preferable [TC07]. The runtime performance heavily depends on the used operators and their time complexity. A generalization of the performance is not possible. We additionally need to distinguish between iterative and parallel problems, because a GPU-accelerated operator can be more suitable than a CPU-based operator for some problems and vice versa. That makes the considerations difficult. But a mandatory process is the detection of game events in the focus region. These need to be updated when a single lens property changes. A simple brute force search algorithm implemented on the GPU was tested in this regard to give at least a performance hint. The runtime between two frames was measured on the GPU for a duration of 10 s using *OpenGL timer queries* [Sh13] while the position of the lens constantly changed. The average time taken to test 750000 initial game events on a moderate desktop computer[2] is 12.143 ms using a cuboid-shaped lens and 12.207 ms using a sphere-shaped lens. This leaves about 50 ms for other operators.

6 Conclusion and Future Work

Computer games provide numerous data in form of game events that are interesting for developers, publishers and players. The visualization of those events is able to reveal insights into facts from the game environment. In this context, the DSRM is an established and appropriate model for the visualization transformation of game events.

The proposed Game Event Lenses provide a controllable opportunity to explore game events. It is possible to focus on different areas of the visualization while a contextual overview is still intact. But there are remaining issues that could be examined in the future, e. g., visual occlusions of information in three dimensional virtual scenes with corresponding lenses as seen in the examples. Information that are displayed inside the focus area but do not belong to it might hinder the recognition of facts. This should be addressed.

The Game Event Lenses base on the DSRM and therefore focus on operators that process data on different stages. The shader based concept is appropriate to simulate the DSRM because an operator can be easily simulated by a shader program. This GPU-based approach is suitable for general purpose computations that deal with parallel problems. That is not tied to game events only. But issues remain. One problem concerns the linkage of several operators. Currently only the programmer of the graphics system is in total control of a sequence of operators. Well defined interfaces and interaction mechanisms are needed if an ordinary user should be able to design and use its own sequences.

Nevertheless, the lens examples of the prototypical implementation demonstrate the aptitude of the presented idea for at least basic or simple visualizations to start with the visual exploration of game events.

[2] Intel i7-960 @ 3.20 GHz x8, 24 GB RAM, NVIDIA GeForce GTX 480 with 1536 MB RAM, NVIDIA driver version 343.22 on Arch Linux 64 bit with kernel version 3.17.1

References

[BD10] Bruns, R.; Dunkel, J.: Event-Driven Architecture: Softwarearchitektur für ereignisgesteuerte Geschäftsprozesse. Springer-Verlag, Berlin, Heidelberg, Germany, 2010.

[Be11] Bertin, J.: Semiology of Graphics: Diagrams, Networks, Maps. Esri Press, Redlands, CA, USA, 2011. Translated by Berg, W. J. from: Jacques Bertin. Sémiologie graphique: Les diagrammes, Les réseaux, Les cartes. Gauthier-Villars, Paris, France, 1967.

[CR98] Chi, E. H.-H.; Riedl, J.: An Operator Interaction Framework for Visualization Systems. In: Symposium on Information Visualization, InfoVis 1998, Research Triangle Park, NC, USA, Proceedings. IEEE Computer Society, pp. 63–70, 1998.

[EDC13] El-Nasr, M. S.; Drachen, A.; Canossa, A., eds. Game Analytics: Maximizing the Value of Player Data. Springer-Verlag, London, UK, 2013.

[HKP12] Han, J.; Kamber, M.; Pei, J.: Data mining: Concepts and Techniques. Morgan Kaufmann Publishers, Waltham, MA, USA, 3rd edition, 2012.

[JMG02] Jankun-Kelly, T. J.; Ma, K.-L.; Gertz, M.: A Model for the Visualization Exploration Process. In (Moorhead, R.; Gross, M.; Kenneth, I. J., eds): IEEE Visualization 2002, Boston, Massachusetts, USA, Proceedings. pp. 323–330, 2002.

[Ko04] Koster, R.: A Theory of Fun for Game Design. Paraglyph Press, Phoenix, AZ, USA, 2004.

[Ma04] MacEachren, A. M.: How Maps Work: Representation, Visualization, and Design. The Guilford Press, New York, NY, USA, 2004.

[Ma13] Mandl, P.: Grundkurs Betriebssysteme: Architekturen, Betriebsmittelverwaltung, Synchronisation, Prozesskommunikation. Springer Vieweg, Wiesbaden, Germany, 3rd edition, 2013.

[Ow05] Owens, J. D.; Luebke, D.; Govindaraju, N.; Harris, M.; Krger, J.; Lefohn, A.; Purcell, T. J.: A Survey of General-Purpose Computation on Graphics Hardware. In: Eurographics 2005: State of the Art Reports. pp. 21–51, 2005.

[Sh13] Shreiner, D.; Sellers, G.; Kessenich, J.; Licea-Kane, B.: OpenGL Programming Guide: The Official Guide to Learning OpenGL, Version 4.3. Addison-Wesley, 8th edition, 2013.

[Sp01] Spence, Robert: Information Visualization. Addison-Wesley, Harlow, England, 2001.

[TC07] Thropp, J. E.; Chen, J. Y. C.: Review of Low Frame Rate Effects on Human Performance. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 37(6):1063–1076, 2007.

[TFS08] Thiede, C.; Fuchs, G.; Schumann, H.: Smart Lenses. In (Butz, A.; Fischer, B.; Krüger, A.; Oliver, P.; Christie, M., eds): Smart Graphics: 9th International Symposium on Smart Graphics, SG 2008, Rennes, France, Proceedings. Springer-Verlag, Berlin Heidelberg, Germany, pp. 178–189, 2008.

[To11] Tominski, C.: Event-Based Concepts for User-Driven Visualization. Information Visualization, 10(1):65–81, 2011.

[To14] Tominski, C.; Gladisch, S.; Kister, U.; Dachselt, R.; Schumann, H.: A Survey on Interactive Lenses in Visualization. In (Borgo, R.; R., Maciejewski; Viola, I., eds): EuroVis - STARs. The Eurographics Association, pp. 43–62, 2014.

[Wa04] Ware, C.: Information Visualization: Perception for Design. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2nd edition, 2004.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Entwicklung eines Spieles zum Thema Schatten-IT

Jan Asmuth[1], Markus Bleß[2], Doreen Brunner[3], Mark Deppe[4] und Marcel Prügel[5]

Abstract: Schatten-IT sind Systeme, Organisationsstrukturen oder Prozesse, die Geschäftsprozesse unterstützen, aber weder technisch noch strategisch ins IT-Service-Management integriert sind. Um in Unternehmen ein Bewusstsein für Schatten-IT zu schaffen, wurde ein Lernspiel entwickelt, welches die Facetten von Schatten-IT spielerisch verdeutlichen soll. In diesem Paper wird der Prototyp dieses Spieles vorgestellt sowie eine Auswertung des Testspielens geliefert. Mit Experten auf dem Gebiet der Schatten-IT und mit Mitarbeitern aus Fach- und IT-Abteilungen wurde der Prototyp des Spieles hierfür zuvor getestet.

Keywords: Schatten-IT, IT-Strategie, Schatten-IT Spiel, Lernspiele, Gamification, Serious Games

1 Einleitung

Schatten-IT ist ein weit verbreitetes Phänomen mit verschiedenen Erscheinungsformen, und Auswirkungen [Ra05]. Unternehmen fehlt heute oft das Bewusstsein für die Arten, Risiken und Chancen von Schatten-IT [ZR12]. Um ein besseres Bewusstsein zu schaffen, wurde im Rahmen eines studentischen Projektes ein Brettspiel entwickelt, welches die Risiken und Chancen von Schatten-IT anhand von Beispielen aufzeigt. Ein höheres Bewusstsein kann beim Management von Schatten-IT helfen. Der Ansatz, Betroffene spielerisch über Schatten-IT aufmerksam zu machen oder aufzuklären, ist in der Literatur nicht zu finden und wäre auch in der Praxis neu. Das Prinzip, Lerninhalte in Form eines Spieles zu vermitteln, hat den Vorteil, dass die Lernmotivation meist höher ist als bei der traditionellen Wissensvermittlung, in Form von Vorträgen oder Infobroschüren [RW15]. Beim Spielen konkurrieren zwei Teams aus jeweils zwei oder drei Spielern – das soll die Kooperation fördern, deren Mangel oft Haupttreiber für Schatten-IT ist [ZR12], [Re12]. Ziel dieses Projektes war es ebenfalls die empfundenen Auswirkungen auf Schatten-IT-Experten und Laien aus Fachbereichen und IT-Abteilungen zu testen.

Mithilfe des Spieles sollen die Mitarbeiter aus Fachabteilungen lernen, weshalb der

[1] HTWG Konstanz, Fakultät Informatik, Zeppelinstraße 16b, 8280 Kreuzlingen, Schweiz, jaasmuth@htwg-konstanz.de
[2] HTWG Konstanz, Fakultät Informatik, Zeppelinstraße 16b, 8280 Kreuzlingen, Schweiz, mabless@htwg-konstanz.de
[3] HTWG Konstanz, Fakultät Informatik, Schulthaißstraße 1, 78462 Konstanz, dobrunne@htwg-konstanz.de
[4] HTWG Konstanz, Fakultät Informatik, Am oberen Schlossberg 9, 71686 Remseck am Neckar, madeppe@htwg-konstanz.de
[5] HTWG Konstanz, Fakultät Informatik, Herrenstraße 36, 76133 Karlsruhe, mapruege@htwg-konstanz.de

Einsatz von Schatten-IT der IT-Abteilung Probleme bereiten kann. Gleichzeitig soll die IT-Abteilung nachvollziehen können, wieso Mitarbeiter aus Fachabteilungen überhaupt Schatten-IT nutzen [ZR12], [Jo04]. Außerdem soll vermittelt werden, dass Schatten-IT nicht nur ein Risiko darstellt, sondern auch eine Chance bedeuten kann, beispielsweise aufgrund des innovativen Charakters der implementierten Lösung. Im Spielverlauf sollen die Spieler Projekte durchführen, bei denen sie selbstständig entscheiden können, ob sie Schatten-IT oder eine offizielle Lösung der IT-Abteilung (formale Lösung) verwenden. Die Projekte sind beispielhaft beschriebene Schatten-IT-Szenarien, die auf die eigenen Erfahrungen der Spieler übertragbar sind.

In diesem Paper werden zunächst die Arten, Chancen und Risiken von Schatten-IT erläutert. Anschließend werden Grundlagen über Lernspiele genannt. Dies soll verdeutlichen, weshalb die Entwicklung eines Lernspieles für die Vermittlung von Wissen über Schatten-IT in Unternehmen sinnvoll erscheint. Hiernach folgen die Vorstellung des Prototyps des Schatten-IT-Spiels, sowie erste Erfahrungen mit Testspielen. Am Ende werden die Erkenntnisse zusammengefasst und ein Ausblick für die Weiterentwicklung des Spieles geliefert.

2 Schatten-IT

Schatten-IT bezeichnet alle Anwendungen, die ohne IT beschafft und ohne Genehmigung des Unternehmens betrieben werden [Ma14], [RZ15], [ZR12]. Das schließt IT-Systeme, IT-Serviceprozesse und IT-Mitarbeiter ein, die von Fachabteilungen eigenständig eingesetzt werden [Ku13], [ZR12]. Schatten-IT ist weder technisch noch strategisch in das IT-Service-Management eingebunden. Synonyme Bezeichnungen für Schatten-IT sind „graue IT“, „Schlingel-IT“ (engl. „rogue IT“) oder „versteckte IT“ (engl. „hidden IT“) [ZR12].

2.1 Schatten-IT in der Praxis

Der Einsatz von Schatten-IT ist weitreichend in der Praxis verbreitet [Ra05]. In einer Umfrage von Frost & Sullivan im Auftrag von McAfee wurden je 300 Mitarbeiter aus IT und Fachabteilungen aus Unternehmen mit über 1000 Mitarbeitern aus Amerika, England, Australien und Neuseeland zum Einsatz von Schatten-IT befragt. Die befragten Mitarbeiter gaben an, dass sie Entscheidungen bezüglich des Zukaufs von Software entweder treffen oder beeinflussen. Nach dieser Umfrage nutzen mehr als 80% der Mitarbeiter nicht genehmigte Anwendungen, die über das Internet vertrieben werden. Diese Anwendungen machen beinahe 35% dieser Anwendungen der Unternehmen aus [Th13]. Ebenfalls können komplexere Lösungen autonom programmiert oder extern zugekauft und durch die Fachabteilungen koordiniert werden, ohne die IT-Abteilung zu informieren oder zu involvieren [ZR12]. Das Vorgehen beschränkt sich jedoch nicht nur auf die Fachabteilungen, denn 83% der Befragten aus der IT nutzen ebenfalls nicht autorisierte IT-Systeme. Aus diesen 83% gaben 39% der IT-Fachleute an, dass sie

Schatten-IT nutzen, um „damit IT-Abläufe zu umgehen“ und 18% gaben an, dass IT-Restriktionen „ihnen die Arbeit erschweren“ [Th13].

2.2 Entstehung von Schatten-IT

Schatten-IT entsteht, wenn die Abstimmung zwischen IT und Fachbereichen schlecht ist oder Schatten-IT einen Mehrwert stiftet, den die IT-Abteilung nicht liefert [Ma14], [RZ15], [ZR12].

Typischerweise entsteht dann Schatten-IT, wenn die Organisationsstruktur der IT-Abteilung nicht zu der Organisationsstruktur der übrigen Fachbereiche passt. Ein mögliches Problem ist beispielsweise, wenn Mitarbeiter aus den Fachabteilungen bei Bedarf für neue IT-Lösungen nicht wissen, wen sie in der IT-Abteilung ansprechen sollen. In diesem Szenario werden häufig IT-Lösungen gesucht ohne die IT-Abteilung miteinzubeziehen [ZR12]. Andererseits kann eine zu rigide Organisation der IT-Serviceprozesse dazu führen, dass Schatten-IT-Anwendungen aus Notwehr von Mitarbeitern aus den Fachabteilungen eingeführt werden, um Flexibilität zu wahren [Ma14], [ZR12], [Wa13]. Ein weiterer Entstehungsgrund von Schatten-IT, der einer schlechten Abstimmung von IT und Fachbereichen zugeordnet wird, ist ein ungenügender Koordinationsapparat für IT-Leistungen. Beispielsweise würde eine fehlende Transparenz bei internen IT-Verrechnungspreisen dazu führen, dass Mitarbeiter aus Fachabteilungen besser kalkulierbare Kosten durch Schatten-IT-Lösungen suchen. Verzögerungen von angeforderten IT-Serviceangeboten, beispielsweise verursacht durch finanzielle oder personelle Kapazitätsengpässe in der IT-Abteilung, können ebenfalls Schatten-IT verursachen. Hierbei handeln Mitarbeiter der Fachabteilungen aus Notwehr, um ihre Flexibilität zu sichern [Be09], [ZR14]. Treiber für Schatten-IT kann schlichtweg die Gewohnheit der Mitarbeiter der Fachbereiche sein, eigenständig Entscheidungen zu treffen, sodass IT-Lösungen zur Prozessunterstützung selbst entwickelt oder beschafft und autonom betrieben werden. Unter Umständen geschieht hier sogar eine Abwägung bezüglich erwarteten Kosten und Nutzen von Schatten-IT-Lösung und formeller IT-Lösung ohne Berücksichtigung der strategischen Vor- und Nachteile oder der direkten Risiken [ZR14]. Weitere Gründe für die Entstehung von Schatten-IT können eine räumliche Verteilung oder fehlender Bezug zur IT-Abteilung sein. Räumlich verteilte Organisationen schränken Mitarbeiter der Fachabteilungen beim IT-Support durch die IT-Abteilung oft ein, sodass die Bereitschaft zur Umsetzung von Schatten-IT-Lösungen gefördert wird. Fehlt ein Bezug zur IT-Abteilung, beispielsweise durch eine geringe organisatorische Verbundenheit, bedingt durch eine Merger & Akquisition-Historie, sind Mitarbeiter der Fachbereiche eher dazu geneigt Schatten-IT-Lösungen umzusetzen [RZ15], [ZR12].

2.3 Folgen von Schatten-IT

Schatten-IT generiert einerseits Mehrwert durch Chancen und Innovation und birgt andererseits Risiken und Konfliktpotential [RZ15].

Schatten-IT-Lösungen wohnt ein starker Geschäftsprozessbezug inne, da sie direkt aus geschäftlichen Problemstellungen von den Anwendern abgeleitet werden. Dadurch können schnell innovative Ideen umgesetzt werden. Ebenfalls fallen die Barrieren der formellen Abstimmungswege, sodass Schatten-IT-Anwendungen flexibel und anpassbar sind. Auch wirken Schatten-IT-Lösungen aufgrund der kreativen und kollaborativen Entwicklung und durch eine hohe Identifikation mit der eigenen Lösung oftmals motivierend [ZR12].

Auf der anderen Seite mangelt es der Schatten-IT an Professionalisierung in Bezug auf Integration, Sicherheit und Datenschutz, sowie Support und Pflege [Gy12]. Gleichzeitig droht die fehlende Professionalisierung der Schatten-IT eine IT-Architektur zu fördern, die schwer zu pflegen und auszubauen ist. Die fehlende Integration kann zu Silo-Lösungen in den Fachbereichen und Ineffizienzen führen und steigert die Tendenz zu einem redundanten und heterogenen Datenbestand. Mangelnder Schutz der Daten verletzt häufig die Compliance und macht das Unternehmen anfälliger gegenüber Spionage. Durch fehlenden Support und Pflege können Folgekosten entstehen. Beispielsweise müssen Fehler und Ausfälle eigenständig gelöst werden und Änderungen in anderen Systemen können dazu führen, dass die Schatten-IT-Anwendung angepasst werden muss. Dadurch können Mitarbeiter des Fachbereichs nicht ihre Geschäftsprozesse ausführen, bis die Schatten-IT-Anwendung wieder betrieben werden kann [Ma14]. Die Zusatzarbeit für das Betreiben von Schatten-IT-Anwendungen senkt insgesamt die Performance des Unternehmens [RZ15]. Entstehende Transparenzverluste behindern die Weiterentwicklung der IT-Landschaft, wodurch die IT des Unternehmens Flexibilität einbüßt [Ma14].

3 Lernspiele in Unternehmen

Lernspiele oder auch didaktische Spiele sind Spiele, die ein implizites Lernen im Rahmen eines spielerischen Kontexts ermöglichen. Bei dem Lerninhalt eines Lernspiels kann es sich um Fähigkeiten, Kulturtechniken oder Wissen zu bestimmten Themen handeln. [WR14] Eine spielerische Vermittlung von Wissen hat klare Vorteile gegenüber den traditionellen Weiterbildungsmethoden in Unternehmen. Die Vorteile liegen beispielsweise in der künstlichen Spielumgebung, die es ermöglicht Zusammenhänge zu vereinfachen und die Bedeutung formaler Hierarchien abzuschwächen. [Sp14] Heutzutage werden diese Vorteile bereits in Lernspielen zum Training von neuen Mitarbeitern, Sachbearbeitern oder Führungskräften genutzt. Führungskräfte können mithilfe solcher Lernspiele beispielsweise ihr Wissen im Bereich Projektmanagement erweitern. Um solches Fachwissen, aber auch z.B. Fähigkeiten im

Bereich Soft Skills zu vermitteln, eignen sich besonders Lernspiele nach dem Schema eines „Serious Games“. [Go11]

3.1 Serious Games

Ein Serious Game ist eine Art von Lernspiel, bei dem Wissen und Informationen zu einem bestimmten Fachgebiet vermittelt werden. Serious Games können beispielsweise innerhalb von Unternehmen genutzt werden um Verkaufstechniken oder Führungsqualitäten zu erlernen oder ein Bewusstsein für bestimmte Sicherheitsrisiken zu schaffen (auch Awareness Games genannt). [Ka12], [Sp14] Serious Games können als Digitale Spiele, Brettspiele oder mit einer Mischung von digitalen und haptischen Elementen realisiert werden. Gerade Serious Games in Form von Brettspielen lassen sich gut im Rahmen eines unternehmensinternen Workshops einsetzen, da außer dem Brettspiel weiter nichts benötigt wird und die Spieler sich gemeinsam auf einen Gegenstand konzentrieren.

3.2 Brettspiele

Wie die technische Umsetzung eines Spiels auf einen Spieler wirkt, hängt von dem jeweiligen Lerntyp ab, der sich von Mensch zu Mensch unterscheidet. Nach der Theorie von Vester gibt es vier Lerntypen: Auditiv, Optisch/Visuell, Haptisch und Intellekt. [Ve95 S. 49-52] Brettspiele sprechen insbesondere Personen an, die zur Gruppe des haptischen Lerntyps gehören. Zwar können konkrete Handlungen aus der Realität auch in Brettspielen oft nur abstrakt durchgeführt werden, jedoch ermöglicht das Erfahren der haptisch verfügbaren Ressourcen und Elemente einen nachhaltigen Lerneffekt für diesen Lerntyp. Der auditive Lerntyp wird bei Brettspielen durch das Vorlesen von Aktions- oder Aufgabenkarten angesprochen. Auch die optische Umsetzung eines Brettspiels kann bereits Informationen und Wissen beinhalten, dies kommt gerade dem visuellen Lerntyp zugute. Ein Brettspiel, das eine gute Haptik und Usability aufweist, kann die Erfahrungen und Eindrücke der Spieler verstärken. Jedoch kann die physische Ausführung eines Brettspiels nicht über Schwächen in der Spielkonzeption, wie fehlende Ziele oder unlogische Regeln hinwegtäuschen [Sp14].

3.3 Konzeption

Der Spielspaß bzw. die Motivation und die Akzeptanz eines Spiels hängen maßgeblich von einzelnen Elementen innerhalb des Spielkonzeptes ab. Die essentiellen Elemente, die bei einer Spielkonzeption beachtet werden sollten, sind Ziele, Auswahlmöglichkeiten, Fortschritt, Regeln, „Soziale Komponente“, Balance und Freiwilligkeit [Sp14]:

Ein Spiel braucht Ziele, die als roter Faden von Spielbeginn bis Ende Bestand haben. Durch sie wird ein Spannungsbogen geschaffen. Das Ziel eines Spieles sollte mit

möglichst wenig Erklärung verstanden werden, um den Einstieg zu erleichtern und erste positive Erfahrungen zu schaffen. Auswahlmöglichkeiten sind wichtig, damit die Spieler erkennen, dass ihre Entscheidungen einen Einfluss auf den Spielverlauf haben. Für jeden sollte der aktuelle Fortschritt des Spiels erkennbar sein, damit er zum Weiterspielen motiviert ist. Nur, wenn der Fortschritt zur nächsten Stufe klar ist, kann der Spieler logische Entscheidungen, im Hinblick auf die Ziele des Spiels, treffen. Regeln strukturieren den Spielablauf und sind einer der wichtigsten Aspekte, wenn es um die Akzeptanz eines Spiels geht. Regeln müssen daher nachvollziehbar, eindeutig und im Vorhinein festgelegt sein. Die „Soziale Komponente“ eines Spiels beschreibt die Anreize und Möglichkeiten der Spieler zur Interaktion während des Spiels. Erst durch die “Soziale Komponente“ wird das Spielen zum Gemeinschaftserlebnis. Eine der wichtigsten Voraussetzungen für den Spielspaß und die Akzeptanz eines Spiels ist die richtige Spielbalance. Aufgaben und Entscheidungen sollten anspruchsvoll, jedoch nicht zu schwer sein. Sowohl zu viele als auch zu wenige Auswahlmöglichkeiten können ein Grund für Frustration sein. Auch die Anzahl an zu vergebenden Punkten und Ressourcen sollten so gewählt werden, dass das Spiel nachvollziehbar und realistisch ist. Die Freiwilligkeit kann als eine Art Hygienefaktor betrachtet werden. Ist die Teilnahme am Spiel nicht freiwillig, wird das Spiel als lästiger Zwang empfunden, der nur schwer durch die motivierenden Anreize des Spiels überwunden werden kann. [Sp14], [BH05]

4 Schatten-IT-Brettspiel

Der Ansatz dieser Arbeit ist es, den Problemen von Schatten-IT durch ein Brettspiel zu begegnen. An dieser Stelle ist das vorrangige Ziel den Spielern zu zeigen, dass Unterschiede zwischen Schatten-IT und formaler IT bestehen. In diesem Zusammenhang soll Verständnis für die Entstehung von Schatten-IT, für die Risiken und Chancen, sowie für die Relevanz der Kooperation Fachabteilungen und IT-Abteilung geschaffen werden. Brettspiele sind dafür geeignet diese Anforderungen zu erfüllen, da Kooperation der Spieler aus verschiedenen Abteilungen oder Hierarchieebenen gefördert wird. In diesem Kapitel werden das entwickelte Schatten-IT-Brettspiel erklärt und zuvor dessen Konzeption vorgestellt.

4.1 Konzeption

An dieser Stelle wird auf die konzeptionellen Aspekte des Schatten-IT-Brettspiels „Ziele“, „Auswahlmöglichkeiten“, „Fortschritt“, „Regeln“, „Soziale Komponente“, „Balance“ und „Freiwilligkeit“ eingegangen.

Ziel des Spiels ist es, als Team Punkte zu sammeln, bevor das konkurrierende Team den zum Sieg benötigten Punktestand erreicht. Dabei stehen jedem Spieler Auswahlmöglichkeiten zur Verfügung, die den Spielverlauf prägen. Innerhalb des Schatten-IT-Brettspiels muss sich jeder Spieler, abhängig von der aktuellen Spielsituation, entscheiden ob Punkte mittels Schatten-IT oder formaler IT erarbeitet

werden. Der Fortschritt des Spiels ist für jeden Spieler transparent, da alle Aktionen und Karten offen auf dem Tisch liegen. Innerhalb des Spiels können Situationen entstehen, bei denen der eigene Fortschritt von anderen Spielern abhängt, sodass logische Entscheidungen oftmals besser über Absprache getroffen werden müssen. Die Regeln müssen im Vorfeld kurz erklärt werden, wobei die meisten Regeln intuitiv gestaltet sind oder bekannte Spielmechanismen beinhalten. Die soziale Komponente findet im Schatten-IT-Brettspiel besondere Beachtung, um die Kooperation zwischen IT-Mitarbeitern und Mitarbeitern aus den Fachabteilungen zu fördern. Das Team muss sich untereinander und sogar mit gegnerischen Spielern absprechen oder muss beobachten welche Ziele welcher Spieler wahrscheinlich verfolgt, um selbst logische Entscheidungen treffen zu können. Das Schatten-IT-Brettspiel auszubalancieren hat eine besondere Herausforderung dargestellt. Einerseits muss das Verhältnis zwischen den Siegpunkten und dem Risiko, Punkte zu verlieren, stimmen. Andererseits muss das Verhältnis der Schwierigkeit einer Aufgabe zu potentiellen Punkten stimmen. Weiterhin ist auch das Verhältnis von Attraktivität von Schatten-IT-Lösung zu formaler IT-Lösung zu berücksichtigen. Die Freiwilligkeit zur Spielteilnahme können wir innerhalb des Spiels nicht gewährleisten, da die intrinsische Motivation das Schatten-IT-Brettspiel zu spielen nicht im Vorfeld gefördert werden kann.

4.2 Das Spielfeld

Das Spielfeld besteht aus einem „Projekte“-Feld, dem blauen „IT-Abteilung“-Feld und den vier gelben Fachabteilungs-Feldern „Marketing“, „Produktion“, „Finanzen“ und „Forschung & Entwicklung“. Im Feld „Projekte“ starten alle Spieler, die anderen Felder sind Zielfelder, die im Laufe des Spiels angesteuert werden können. Alle Felder dazwischen sind lediglich Weg-Felder und haben keine weitere Bedeutung, eine Ausnahme bilden die „Kaffeeecke“-Felder.

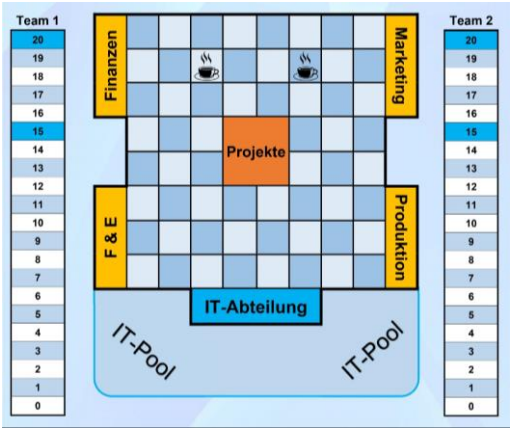


Abb. 1: Das Spielfeld

Unterhalb der IT-Abteilung befindet sich der IT-Pool, hier werden die IT-Chips gelagert, bis sie von einem Spieler abgeholt werden. Diese werden später benötigt um Projekte umzusetzen und simulieren die IT-Kosten (Geld, Zeit und andere Ressourcen) eines Projekts. Da es nur einen begrenzten Pool an Chips gibt, sind diese oftmals aufgebraucht. Das soll die Auslastung der IT-Abteilung simulieren und bildet einen oft auftretenden Grund für den Einsatz von Schatten-IT: Notwehr. Wenn Mitarbeiter nicht die Möglichkeit haben, die IT-Abteilung mit einzubeziehen, da diese ausgelastet ist, nötigt sie dieser Umstand dazu, selbst Schatten-IT zu generieren. Rechts und links neben dem Spielfeld befinden sich die Siegpunkte-Skalen der beiden Teams. Diese geben den aktuellen Punktestand des jeweiligen Teams an. Das sogenannte Risiko-Rad gibt an welches Risiko eintritt.

4.3 Der Spielverlauf

Das Spiel kann mit vier oder sechs Personen gespielt werden. Diese spielen jeweils zu zweit, bzw. zu dritt in einem Team zusammen. Jeder Spieler bekommt jedoch eine eigene Spielfigur und darf diese frei auf dem Spielfeld bewegen. Die Menge an Feldern, die eine Figur bewegt werden darf, wird durch einen einzelnen Würfel vorgegeben. Eine Ausnahme bildet die Augenzahl „1“. Wird diese gewürfelt, muss der Würfelnde eine Aktionskarte ziehen, die im Kapitel „4.5 Die Sonderkarten“ erläutert wird. Immer wenn ein Spieler das Projekte-Feld betritt (auch zu Beginn des Spiels), darf er eine Projekt-Karte ziehen. Dieses Projekt kann er dann entweder als Schatten-IT Lösung sofort umsetzen oder mit Unterstützung der IT-Abteilung. Entscheidet er sich für die zweite Alternative, muss er zunächst die benötigten IT-Chips aus der IT-Abteilung besorgen. In beiden Fällen muss er aber in die jeweilige Fachabteilung des Projektes wandern um dieses umzusetzen. Allerdings ist der direkte Weg zur Fachabteilung natürlich deutlich kürzer. Das soll den Aufwand simulieren, den ein Mitarbeiter hat, wenn er die IT-Abteilung konsultiert. Dieser Schritt ist oftmals mit einigen Schwierigkeiten verbunden, da entweder Formulare ausgefüllt werden müssen oder der Aufwand, der IT-Abteilung das Problem zu erläutern, recht groß ist. Dadurch kommen Mitarbeiter in Versuchung, das Problem einfach selbst anzugehen. Im Spiel muss der Spieler also abwägen, ob sich der Mehraufwand zur IT-Abteilung zu wandern lohnt, oder ob er das Risiko einer Schatten-IT Lösung eingeht. Sammelt ein Spieler viele Schatten-IT-Lösungen, steigt dessen Risiko Punkte zu verlieren. Das kann einerseits durch eine höhere Streuung des Risikos der Fall sein, andererseits könnten mehrere Projekte gleichzeitig scheitern. Dadurch bekommt der Spieler ein Gefühl für das Risiko von Schatten-IT. War jeder Spieler an der Reihe wird am Ende jeder Runde das Risikorad gedreht.

4.4 Die Projekte

Auf der Vorderseite der Projektkarte (in Abbildung 2 links dargestellt) befinden sich eine kurze Beschreibung des Szenarios sowie die Lösung der IT-Abteilung. Diese kostet IT-Chips, die zunächst von der IT-Abteilung abgeholt werden müssen. Anschließend

bringt der Spieler das Projekt in die Fachabteilung und kann es dort gegen die angegebene Menge an Siegpunkten einlösen.

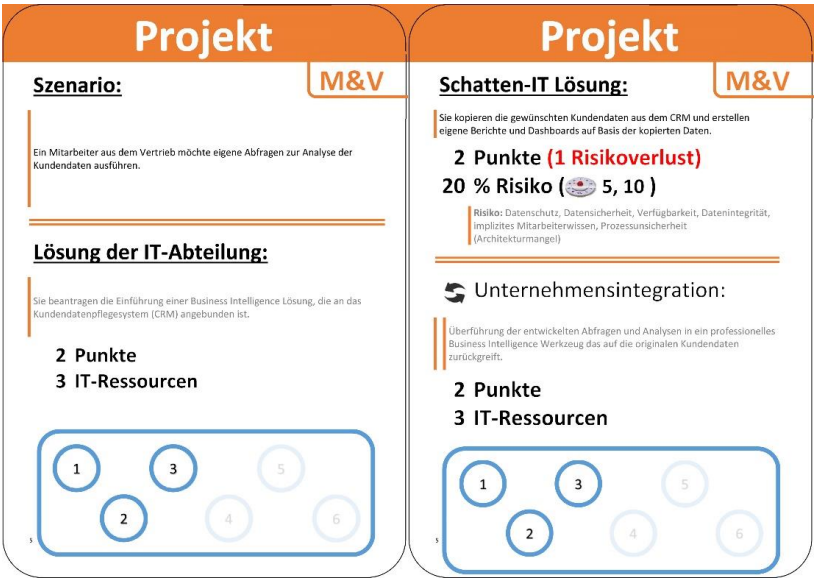


Abb. 2: Die Projektkarte

Alternativ kann auch ohne Umweg in die Fachabteilung gewandert werden. In diesem Fall wird die Schatten-IT Lösung auf der Rückseite gewählt. Diese Lösung gibt zwar genauso viele Punkte wie die Lösung der IT-Abteilung, ist aber mit einem gewissen Risiko behaftet. Wie bereits erwähnt wird nach jeder Runde das Risikorad gedreht. Wird eine der auf der Projektkarte in Klammer geschriebenen Zahlen „erdreht“, verliert das Team die in Rot angegebene Menge an Siegpunkten. Ob sich das Risiko lohnt, muss das Team je nach Szenario und Risikobereitschaft selbst entscheiden. Wird das angesammelte Risiko durch die steigende Anzahl an Schatten-IT Projekten zu groß, kann das Team überlegen, Schatten-IT Projekte in das Unternehmen zu integrieren und so das Risiko zu vermeiden. Dadurch können zusätzliche Siegpunkte geschaffen (siehe Abbildung 2 rechts unten), es muss jedoch ein weiterer Umweg zur IT-Abteilung in Kauf genommen werden. Der Grundgedanke einer Schatten-IT-Lösung ist in den meisten Fällen innovativ, benutzernah und vor allem problemorientiert. Die Kosten sind oft geringer und die Selbstidentifikation durch die Anwender höher. Daher ist es in vielen Fällen sinnvoll, die Schatten-IT-Lösung zu professionalisieren und durch die IT-Abteilung in die Unternehmens-Infrastruktur aufzunehmen. Dies lohnt sich jedoch nicht in allen Fällen, daher ist es auch hier Sache des Teams, wie es sich zum jeweiligen Projekt entscheidet. Dadurch soll den Spielern klar gemacht werden, dass Schatten-IT nicht prinzipiell schlecht ist, sondern dass es auch Fälle gibt, in denen Schatten-IT dem Unternehmen einen Mehrwert bringen kann. Trotzdem sollte Schatten-IT überwacht, kontrolliert oder als formale Lösung übernommen werden.

4.5 Die Sonderkarten

Zusätzlich zu den bisher genannten Mechanismen gibt es zwei Typen von Sonderkarten. Diese sollen das Spiel etwas abwechslungsreicher gestalten. Wie bereits zuvor beschrieben, gibt es auf dem Spielplan sogenannte „Kaffeeecke“-Felder. Wenn diese angesteuert werden, darf der Würfelnde eine Kaffee-Karte ziehen. Auf dieser ist ein positiver Effekt für den Spieler, sodass es sich in vielen Fällen lohnt das Feld aufzusuchen, falls die Spielfigur in der Nähe ist. Diese Kaffeeecken-Karten sollen dem Spieler vermitteln, dass es Vorteile bringt, sich außerhalb der Abteilungen untereinander auszutauschen.



Abb. 3: Die Aktions- und Kaffeekarten

Die Aktionskarten dagegen müssen gezogen werden, wenn eine „1“ gewürfelt wurde. Der Würfelnde liest diese allen laut vor. Im Gegensatz zu den Kaffeeecken-Karten können die Aktionskarten sowohl positive als auch negative Effekte beinhalten. Die Aktionskarten befassen sich thematisch ebenfalls mit Schatten-IT und sollen beispielsweise Notwehr-Gedanken bei den Spielern auslösen.

4.6 Erfahrungen mit dem Schatten-IT Spiel

Das Schatten-IT Spiel wurde mit 18 Testspielern in vier Runden mit dreimal vier und einmal sechs Spielern getestet. Die Testspieler wurden auf Basis der verschiedenen Zielgruppen des Spiels ausgesucht. So waren Vertreter des Rechenzentrums der Fachhochschule, als auch Professoren und Studenten mit und ohne Fachwissen Teil des Probespiels. Zudem wurde ein Probespiel in den Teambildungstag einer ERP-Beratungsfirma integriert, bei welchem Fachbereich und IT anwesend waren. Bei diesen Probespielen wurde zuerst das Spiel und dessen Regeln durch einen Spielleiter erklärt.

In den Spielverlauf wurde nur eingegriffen, wenn sich eine akute Frage ergab oder eine Regel nicht beachtet wurde. Nach dem Spielen wurde den Spielern ein Fragebogen ausgeteilt. Dieser sollte den subjektiven Erkenntnisgewinn zum Thema Schatten-IT ermitteln. Des Weiteren gab es Fragen zum Thema Spielspaß, Logik und Verständlichkeit des Spiels.

Alle 15 Testspieler hatten Spaß beim Abarbeiten der Szenarien. Die Logik und die Szenarien des Spieles waren verständlich und wurden mit den Noten Ø 1,6 und Ø 1,4 auf einer Skala von 1 (sehr gut) bis 5 (sehr schlecht) beurteilt. Der Lernerfolg wurde positiv bewertet (Ø 2,1), einzig bei der Fragen „Sind Ihnen die Chancen von Schatten-IT durch das Spiel bewusst geworden?“ erhielten wir durchwachsene Kritik von Ø 2,64 (Median 3). Die Ursachen von Schatten-IT und deren Verdeutlichung im Spiel durch die begrenzte Anzahl an IT-Chips wurden jedoch wieder mit Ø 1,7 positiv bewertet.

Im anschließenden Gespräch mit den Spielern wurde deutlich, dass an den Szenarien und den jeweiligen Lösungsvorschlägen noch gefeilt werden muss, und gängigere Szenarien gefunden werden sollten. Auch wurden einige kleinere Änderungsvorschläge an der Ablauflogik des Spiels diskutiert. Eine deutlichere Repräsentation der Risiken durch Risikogruppen als nur durch Zahlen wurde ebenso diskutiert wie eine übersichtlichere Gestaltung der Szenarienkarten. Diese Vorschläge sollen im Rahmen zukünftiger Forschung erprobt werden.

5 Zusammenfassung und Ausblick

In diesem Beitrag wurde ein Brettspiel zum Thema Schatten-IT erstellt und getestet. Der Fokus des Spiels liegt darauf, das Bewusstsein für das Phänomen spielerisch zu vermitteln. Inhärente Vorteile des Brettspiels sind der informelle Charakter des Spiels und die persönliche Kooperation zwischen den Mitspielern. Die Erfahrungen aus den Testspielen des Brettspiels zeigen, dass dem komplexen Thema Schatten-IT gut mit Spielen begegnet werden kann.

Das entwickelte Schatten-IT-Brettspiel ist auf 4-6 Spieler beschränkt und dauert mit Vorbereitung durchschnittlich 45 Minuten. Die Anzahl der Personen, die über Schatten-IT aufgeklärt werden sollen, ist durch die geringe Spielerzahl und hohe Spieldauer pro Brettspiel beschränkt. Spannend wäre die Beantwortung der Frage, ob digitale Spiele zum Thema Schatten-IT effizienter sind und dabei die fachübergreifende Kooperation gewährleistet werden kann. Außerdem ist ein weiteres Projekt geplant, das erforschen soll, ob und wie das Bewusstsein für Schatten-IT gesteigert werden kann. In diesem Zusammenhang wird untersucht werden, ob physische Spiele einen höheren Durchsatz erzielen können und welche Spiele für welche Entscheider sinnvoll sind.

Danksagung

Ein besonderer Dank gilt Prof. Dr. Christopher Rentrop, Stephan Zimmermann und Melanie Huber, die uns nicht nur bei der Umsetzung des vorgestellten Projekts, sondern auch bei dem Verfassen dieses Beitrages tatkräftig unterstützt haben. Dieser Beitrag entstand im Rahmen eines Teamprojekts im Masterstudiengang Informatik an der HTWG Konstanz.

Literaturverzeichnis

[Be09] Behrens, Sandy: Shadow systems. ACM 52 (2), S. 124–129, 2009

[BH05] Björk, Staffan; Holopainen, Jussi: Games and Design Patterns. In: The Game Design Reader: A Rules of Play Anthology, Salen, Katie; Zimmerman, Eric, MIT Press, S. 410- 437, 2005

[Go11] Goertz, Lutz: Einsatzmöglichkeiten für Serious Games in Unternehmen. In: Personalführung, 2011

[Gy12] György, Andreas et al.: Exploring the shadows. IT Governance approaches to user-driven innovation. In: ECIS 2012 Proceedings (222), 2012

[Jo04] Jones, David et al.: The Rise and Fall of a Shadow System. Lessons for Enterprise System Implementation. In: ACIS 2004 Proceedings: 96, 2004

[Ka12] Kapp, Karl M.: The gamification of learning and instruction, 2012

[Ku13] Kurzlechner, Werner: Gefahr in der Grauzone, http://www.cio.de/a/gefahr-in-der-grauzone,2915593,3, 21.02.2015

[Ma14] Martin, Wolfgang: Die Schattenseite der Wolke: Tipps zum Umgang mit Schatten (-IT), http://www.wolfgang-martin-team.com/pdf/WXM-Best-in-Cloud-2014-10.pdf, 20.02.2015

[Ra05] Raden, Neil: Shedding Light on Shadow IT. Is Excel Running Your Business? http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.14.3351&rep=rep1&type=pdf, 13.03.2015

[RW15] Reiners, Torsten; Wood, Lincoln C.: Gamification in Education and Business. Springer International Publishing, 2015

[RZ15] Rentrop, C.; Zimmermann, S.: Herausforderungen Schatten-IT, http://kips.htwg-konstanz.de/images/PDF-Content/broschuere.pdf, 17.02.2015

[Sp14] MaibornWolff GmbH: Spiel in der IT-Beratung. München, 2014

[Th13] Frost & Sullivan: The Hidden Truth Behind Shadow IT - Six trends impacting your security posture, http://www.mcafee.com/us/resources/reports/rp-six-trends-security.pdf, 20.02.2015

[Ve95] Vester, Frederic: Denken, Lernen, Vergessen. Teil 1-3, 25, München, 1995

[Wa13] Walters, Richard: Bringing IT out of the shadows. In: Network Security 2013 (4), S. 5–11, 2013

[WR14] Warwitz, Siegbert A.; Rudolf, Anita: Spielend lernen - Lernspiele. In: Vom Sinn des Spielens. Reflexionen und Spielideen, Baltmannsweiler, 3. Auflage, S. 82, 2014

[ZR12] Zimmermann, Stephan; Rentrop, Christopher: Schatten-IT In: HMD 49 (6), S. 60–68, 2012

[ZR14] Zimmermann, Stephan; Rentrop, Christopher: On the Emergence of Shadow IT – A Transactions Cost-Based Approach. In: Proceedings of the 22nd European Conference (ECIS), Tel Aviv, Israel, 2014

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Framework for Fuzzing USB stacks with Virtual Machines

Tobias Mueller[1]

Abstract: The Universal Serial Bus (USB) is a widely deployed technology that connects peripheral devices to computer systems. Despite its popularity and the vast number of existing USB enabled devices, assessing security properties of the USB key-components in an automated fashion has not yet been achieved by existing solutions. While malicious USB devices exist, they still require physical actions such as plugging the device in. At present, arbitrary USB behaviour cannot be implemented in software for easy consumption with virtual machines, thus making it hard to test USB stacks, USB drivers or applications using the functionality a driver exposes.

This paper presents a design to write software defined USB devices, using them to automatically fuzz-test key components of a USB enabled system, and to record abnormal behaviour of systems under test.

The design was prototypically implemented and successfully applied to a simple kernel driver. The results suggest that USB exposes a considerable attack surface and that real attacks are possible. They also demonstrate that the design developed is capable of uncovering flaws in kernel level drivers as well as user space applications. It is also capable of detecting USB stack fingerprints of various operating systems which can enable an attacker to build a physical USB device which, when plugged in to the victims machine, can be used to launch a targeted attack.

Keywords: security, usb, fuzzing

[1] Universität Hamburg

1 Introduction

The Universal Serial Bus (USB) [Co00] is a widely adopted technology which replaced serial and parallel IO ports in 1996. USB emerged because attaching peripheral devices to a PC was cumbersome and error-prone. This was primarily due to the traditional IO paradigm which mapped the devices into the CPU's IO address space and assigned an interrupt line (IRQ). The design of USB allows attaching, configuring and accessing peripheral devices with low cost and simplicity from the user's perspective. Other benefits include one interface for many devices, automatic configuration, hot pluggability or built-in power supply for the device [AI97]. A modern PC is equipped with USB ports to attach keyboards, mice, cameras, printers, scanners, hard-drives, mobile phones or other devices. Even embedded systems such as printers or mobile phones have the ability to attach devices via USB. Key features of USB include its versatility, its inexpensiveness and support by major operating systems.

When an operating system encounters a device attached via USB, it needs to load the appropriate driver to expose the functionality of the device to the user. Presently, although nowadays some drivers can be written in user-space[3], many drivers still reside in kernel space, not only for legacy reasons but also for performance or convenience purposes [CRK05].

Even though kernel developers are usually very experienced and kernel code is subject to repeated review, varying code quality for USB drivers can be expected for multiple reasons. Not only do many different devices exist which require numerous different drivers, but these devices will also have time-to-market constraints which do not allow for driver code security audits. Examining the history of the Linux kernel source code[4] supports that claim.

If those drivers in kernel space were vulnerable, an attacker could gain elevated privileges because it can be assumed that the kernel runs with the highest privileges on the system.

Several concrete attack scenarios exist. An attacker could manipulate elections if a USB-based voting device [ABS07] was used. Publicly available computers with USB interfaces could be attacked by simply plugging in a malicious device. Such a public computer could be a photo-terminal used to load photos to be printend from a pendrive or a digital camera, a PC in a library or an unsupervised machine in a shop. Even some aircraft have an in-flight entertainment system which allows USB devices to be plugged in to listen to songs on portable music players or view documents on a pendrive [Th06]. But the attacker does not necessarily need to physically attach a malicious device herself. Simply distributing a new and expensive looking device (i.e. a digital camera or music player) will most likely lure the victim to plug the device into her computer and thus allow the attacker gain control over it. It is also possible to run USB over IP networks (using USB/IP [Hi05]) and Wireless USB [Le07] uses radio technology as the transport layer for USB. Thus, physical access to the targeted machine is not necessarily needed.

[3] Using libusb, which provides a high level API to communicate with a USB device

[4] i.e. by executing `git log -grep='overflow'` in the USB driver directory

```
class USBInDeviceDescriptor(Packet):
    name = 'DeviceDescriptor'

    fields_desc = [
        LEShortField('bcdUSB', 0x0200),
        ByteEnumField('bDeviceClass', 0, CLASS_ENUMS),
        ByteEnumField('bDeviceSubClass', 0, SUBCLASS_ENUMS),
        ByteEnumField('bDeviceProtocol', 0, PROTOCOL_ENUMS),
        ByteField('bMaxPacketSize', 64),
        LEXShortEnumField('idVendor', 0xffff, VENDOR_ENUMS),
        LEXShortField('idProduct', 0x1337),
        LEShortField('bcdDevice', 0x2342),
        ByteField('iManufacturer', 0),
        ByteField('iProduct', 0),
        ByteField('iSerialNumber', 0),
        ByteField('bNumConfigurations', 0),
    ]
```

List. 1: Python code representing a device descriptor using Scapy

Given the popularity of USB and the expected privileges, exploiting USB is very attractive to an attacker. Consequently, being able to uncover flaws and to fix them is important. Finding bugs, however, often is a non-trivial task, especially in the case of USB, quite simply because providing arbitrary input for the USB driver is not easily possible today, without a USB development board which allows delivery of arbitrary data.

This work extends an existing approach which concentrated on mutation-based fuzzing of the communication between a physical device and a host [JJ10]. Contrasting that existing work, a method to attach USB devices to a virtual machine and to generate arbitrary packets in software is presented here. As the USB devices are written in software , tests can be run without needing to physically attach real USB hardware. Additionally, tests are conducted using packets with payload produced by “fuzzing”.

The contributions of this paper are:

- software-defined USB devices with fuzzing capabilities,
- a framework for executing automated fuzz tests, and
- reliable detection of operating systems through fingerprinting.

This paper is structured as follows. Section 2 describes fuzzing techniques and why fuzzing is a viable approach to identify vulnerabilities. Section 3 lists the components involved around a USB-enabled system. Section 4 discusses how the components can be tested using the described techniques. Section 5 details how the implementation was performed and results obtained. Section 6 concludes this paper and gives an overview of future work.

2 Fuzzing

This section describes the fuzzing technique that will be used to search for vulnerabilities in USB stacks, drivers and applications.

"Fuzzing" (or "fuzz testing") was coined in the late 80's while trying "*to evaluate the robustness of various UNIX utility programs, given an unpredictable input stream*" [TDM08]. With the development of better tools, such as "american fuzzy lop"[5] or "Address Sanitizer"[6], fuzzing gained popularity as a security method. Fuzzing results are likely to be exploitable, because the input, which is necessary to produce the erroneous state, is generated during the fuzzing process. Fuzzing is also well suited to expose bugs in kernel level drivers as shown by Keil and Kolbitsch in [KK07].

Tradionally, fuzzing is used to conduct blackbox and state-less tests with randomly generated bytes. Blackbox because the fuzzer does not know anything about the tested program. Since knowledge of internals of the program is not required, this general purpose approach is widely applicable. The drawback of this method, however, is its inability to test either specific input fields or complex protocols. State-less because the fuzzer generates bytes independently of each other regardless of the context or semantics of that byte. Thus, if the input is expected to contain a checksum over the randomly generated bytes and that checksum is not properly updated, the program aborts processing and the fuzzer tests only the checksum validation code. Those fuzzers are labelled "dumb".

"Smart" fuzzers take into account properties of the expected input and can thus force the tested program to execute more code paths. Also, specific fields of the expected input can be tested, i.e. the field indicating a length. This allows the generated payload to penetrate deeper into the tested program, thus testing more aspects of the program. This is also known as "schema-based fuzzing" because a known pattern is modified [NN08].

Because the USB protocol is stateful and most packets have a fixed structure, this paper follows such a schema-based approach using Scapy [Bi10a]. Scapy is a framework that is designed to interact with packets on ethernet networks. It is primarily used to craft and manipulate packets on various network layers resulting in a powerful networking framework to rapidly create tools for conducting various tests, "*but instead of dealing with a hundred line C program, you only write 2 lines of Scapy*" [Bi10b]. It also provides a `fuzz()` function to generate packets with random yet appropriate values for the packet's fields. This includes the width in bits of the generated value as well as adjusting modelled dependencies such as checksums or length indicating fields. Listing 1 shows how Scapy can be use for USB packets, instead of network packets.

3 USB Components

This section describes the key components that an operating system needs in order to support USB. Some of these components will later be targets of attacks.

[5] `http://lcamtuf.coredump.cx/afl/`
[6] `https://code.google.com/p/address-sanitizer/`

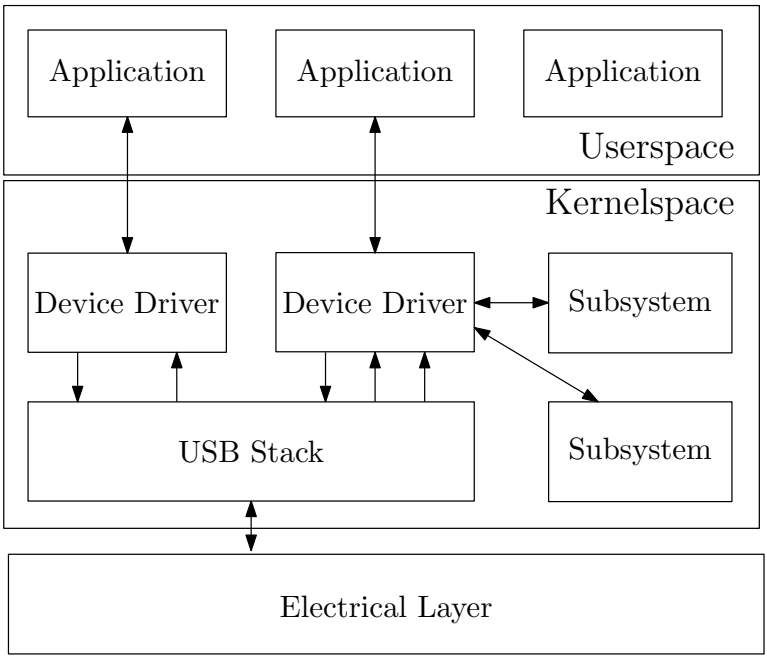


Fig. 1: Key components of a system with USB support

Figure 1 shows the fundamental parts of a USB system. At the bottom a USB device is connected with the USB controller on the host at the electrical layer. The USB controller is controlled by the USB stack which takes care of low level USB message passing and other USB protocol-related aspects such as error correction. The USB stack sends and receives messages from either sources or sinks on the device, called endpoints. Connections to endpoints build logical pipes between the host and the device which are exported to drivers. Upon the USB stack, various drivers are implemented which use the provided USB messaging capabilities to exchange messages either via control, isochronous, bulk or interrupt transfer packets. The drivers may reuse already existing functionality from one or more subsystems (i.e. SCSI for hard-drives or v4l for video cameras). Applications can then use the device through one of the various interfaces that a driver could expose (i.e block device or `mmap()`).

After a device has been plugged in, the operating system asks the device for its details in order to establish the pipes and to know which driver to load. This process is called "enumeration" and it involves so called descriptors to be sent from the device to the host. Every USB device is asked by the USB standard to answer a set of commands that might be issued during enumeration.

As shown in Figure 2, four descriptor types exist: device, interface, endpoint and string descriptor. The very first descriptor requested by the operating system is the device descriptor which describes basic aspects of the device in question such as a globally unique

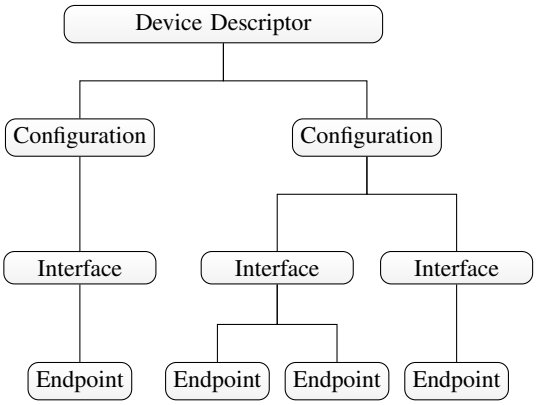


Fig. 2: USB Descriptor hierarchy

vendor ID, a product ID, the "class" of the device and the number of "configurations" which the operating system can select the device to run under. A configuration describes, among other things, whether the device is self-powered or how many "interfaces" the devices exposes. The class indicates that a device speaks a well defined protocol to expose its functionality. Defined classes include mass-storage, audio or video. A device can also announce to not belong to a prefined class or to implement several classes. In the latter case the class information is attached to the interface which is described by an interface descriptor. It also contains information about the number of available endpoints which in turn have their own descriptor defining the packet size and the interval for the host to poll for new messages [Co00].

Once the enumeration is complete, the kernel loads the appropriate driver. To find a suitable driver the kernel first[7] looks at the vendor ID and product ID. If it finds a driver that claims to be responsible for that device it will be loaded. If no driver is found a generic driver for the class of the device (or the interface) will be sought. The driver is then reponsible for exposing features of the device to the user [Ve08].

These components can be attacked using fuzz-testing. The electrical layer, however, is not of interest in this work because we focus on fully automated virtualised testing. Thus physical connections are out of scope. The USB stack can be attacked, i.e. by signalling many device attachments or by interfering with the enumeration process. Knowing the driver-loading mechanism and the structure of the descriptors allows us to trigger a certain driver to be loaded which in turn enables us to send fuzzed messages to the driver and thus assess its robustness. Furthermore, it is possible to send messages to applications if the higher level protocol is known. However, due to their limited applicability and impact, attacks on applications are not as compelling as their driver counterparts. Therefore, focus was given on rendering the latter possible.

[7] This is considering the Linux kernel, but other operating systems do it similarly

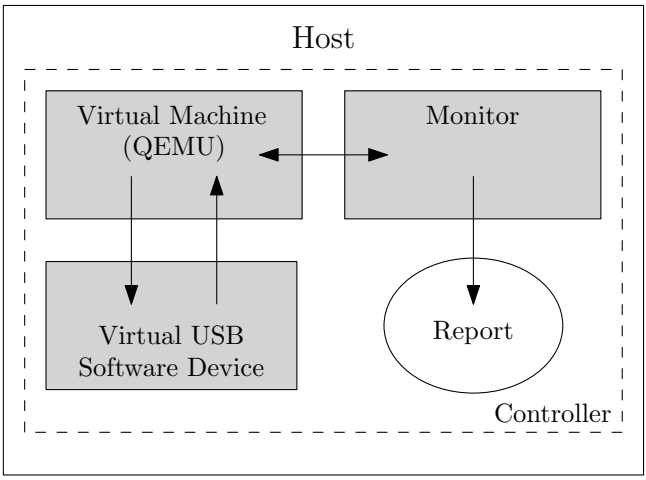


Fig. 3: Architecture of the automated fuzz-testing setup

4 Implementation

To find vulnerabilities in the components mentioned in Section 3 using the techniques described in Section 2, a prototypical architecture was built which allows automated testing. It extends an existing design [JJ10] such that no physical device needs to be connected to the host in order to find vulnerabilities.

Figure 3 depicts the built setup. To allow automated tests the host operating system runs a fully virtualised guest operating system. Note that it is possible to do the testing on the host itself. However, potential crashes of the kernel and lost log-files are inherent drawbacks of that approach. Instead of talking to physical devices, the virtual machine pipes USB communication in and out to a separate process on the host which behaves like a regular USB device. The virtual device then generates USB packets using fuzzing. A monitor watches the guest operating system and reports if unexpected behaviour, i.e. a crash, occurs. A controlling component is responsible for actually spawning the above mentioned components.

The actual implementation uses QEMU [Be05] as virtualisation software to run the guest operating system. QEMU is a free virtualisation solution that fully virtualises different CPUs along with the necessary hardware such as hard-drive, network interface or USB controller. It allows complete and unmodified operating systems to be run in a virtual machine. It also allows the machine state to be saved and loaded at a later time. This can be used to prepare a virtual machine that is fully booted or has a special program started so that testing will take less time. QEMU supports physical USB devices to be passed from the host to the guest. It also supports simple virtual internal USB devices such as keyboard or mouse implemented inside QEMU to allow delivery of keystrokes or mouse movements from the host to the guest.

QEMU was enhanced in two steps: First with a USB packet filter functionality and then with support for external virtual USB devices. The latter is depicted in Figure 3. The USB packet filter intercepts communication between the guest operating system and a USB device attached to a Linux host[8]. The intercepted packets are piped out to a process on the host and read back in. The process on the host can thus either write the packets out to disk or modify them in place.

In order to carry out the schema-based fuzz-testing we need to obtain the schema first. Either valid communication patterns can be crafted by looking at the protocol specification or valid USB packets can be recorded for subsequent decoding. We chose to record valid communication by attaching a physical USB device to the host and thenn pass it through to QEMU. The guest running in QEMU started to communicate with the USB device and the above mentioned USB filter functionality was used to record the USB conversation.

Since the filter allows packets to be modified in place as they are in transit, this would already enable fuzz-testing. However, it would not be automatable since a physical device needed to be plugged in to the host. Also, only dumb fuzzing could be conducted as the obtained packets have not been decoded.

In order to build Scapy models for the packets, the USB specification [Co00] was used to dissect the packets and determine the types of the packet's fields, i.e. `Short` or `IntEnum`. This led to a proper description of the protocol and packets used for enumeration which in turn allowed easy packet generation and modification using Scapy's facilities.

4.1 USB Device Emulation

Although QEMU supports virtual internal USB devices as described above, they are not backed by an external program. To keep this framework modular, a second feature for QEMU was implemented: Support for external software-defined USB devices which are communicated with via a pipe so that USB packets are, as with the USB packet filter mentioned above, piped out to the host and read back in. Because the interface to the virtual machine consists of two named pipes only, the internals of the virtual machine are abstracted and implementing a software USB device is much simpler than writing a new virtual internal USB device for QEMU.

The actual fuzzing is done in the USB Device Emulation component which is responsible for generating USB packets and writing them to the virtual machine's USB in-pipe. Hence, the virtual USB device can do as simple things as replaying already obtained packets. Naturally, dumb-fuzzing can be applied on these packets.

Two generations of software USB devices were produced. The first and simple generation reads previously obtained packets and decodes them. A configuration then tells the software device which fields in which type of packet have to be fuzzed before sending them to the host. Obviously, this can resemble dumb-fuzzing by simply configuring all

[8] Porting this filter to other platforms such as BSD should be trivial

packet types and all fields to be fuzzed. The second generation is a stand-alone state-machine which autonomously answers packets from the host. To save time this automaton did not implement every command required by the USB specification but rather only the commands which were used in the initially obtained USB communication. This allowed for implementation of rudimentary USB devices such as web-cams or hard-drives. Again, packet types and fields to be fuzzed need to be configured.

4.2 Monitoring

In order to detect whether the tested operating system behaves unexpectedly (i.e. crashes), a monitoring component reports back every predefined number of seconds. It needs to implement a reasonable probe routine which is capable of detecting whether the guest operating system is still fully operational. The probe can be implemented as a simple check whether opening a TCP connection to a port on the guest is successful or as a complex routine that logs on to the guest, retrieves and analyses log-files.

In this framework, the monitoring component is realised as a Python module which is loaded in a separate thread by the controlling component. Thus communication between the controlling component and the monitor is possible using Python objects. The separate thread enables the monitor to accidentally block and not stall the execution of the controlling component. The controller regularly executes the monitor's `is_alive` function and assumes the guest operating system to be dysfunctional after a predefined number of probes have not been answered. The monitor is also reponsible for exposing information that is valuable for identifying potential malfunction, i.e. by writing out log files that were obtained from the guest operating system.

4.3 Controlling

A controlling component is responsible for starting the virtual machine, attaching virtual USB devices and alert when the virtual machine does not behave as expected.

To run many fuzz-tests, even on multiple machines, the design presented in this paper does not require elevated privileges and is able to be run in many instances in parallel. Also, the components can easily be exchanged, as long as the interfaces are implemented. The interfaces are simple by design: Two named pipes for the emulated USB device, a Python module for the monitoring instance and a process to be called for the virtual machine. This implies that any type of virtual USB device, not only a fuzzer, can be used in this framework. Given the ability to prototype a USB device in software using a rather high level description for Scapy, this framework might be helpful for driver programmers to test whether their implementation fulfils functional requirements.

5 Evaluation

This section summarises the findings the implementation mentioned in Section 4 yielded.

While testing the second generation emulated USB device with different guest operating systems different behaviour in the enumeration process was observed. Variations are caused by different implementations of the USB stack. Table 1 shows different behaviour

Operating System	**Packet Sequence**	**Retries**	**Remarks**
Windows XP SP 2&3, 7	SETUP, IN, OUT	3	IN length: 64
Linux 2.6.33	SETUP (9 times), RESET	4+2	4 get descr. then 2 set addr.
OpenBSD 4.7	SETUP, IN, OUT	7	IN length: 8
FreeBSD 8.0	SETUP, IN, OUT	6	sets address right away

Tab. 1: USB Stack Fingerprints of various operating systems

of various USB stacks. These samples were obtained by attaching a software USB device that answers with zero bytes only. It it thus possible for a USB device to tell operating systems apart. This enables an attacker to launch platform specific attacks.

It was also observed, that QEMU's virtual USB controller sent uninitialized memory to the device with "IN" transactions, which is similar to Etherleaking which exposes memory via padding for Ethernet packets [Bi10b]. Even if this "USB Leaking" does not happen with real hardware USB controllers it can be used by a USB device to be able to detect whether it is attached to a virtual machine.

The USB stack of the guest operating systems were tested rather by accident than on purpose. While testing whether the implemented USB software devices work, they often crashed and disrupted the communication and produced errors in the guest's USB stack. All errors were handled gracefully by generating a proper error message and thus no flaws were exposed. Trying to rapidly attach many USB devices only uncovered bugs in QEMU and not in the USB stack of the guest operating system. In fact, QEMU monopolizes the CPU after attaching 40 devices to the guest.

To test whether USB drivers could be exploited, a Linux driver with a buffer overflow vulnerability was dedicatedly written. The framework was configured to make the guest load the driver. Upon receiving a specially formatted message, the driver crashed the kernel and left the system in an unoperational state. The monitor realised the machine malfunctioning and reported it to be dead. The controller raised an alarm and made necessary informaton available that is needed to reproduce the crash. Existing drivers, however, were not tested due to time constraints.

The previous paragraphs show that this prototypical implemenation is capable of finding flaws in the identified key components of a USB enabled system. Future work should concentrate on making the presented framework test existing drivers of different operating systems.

6 Conclusion

This section will conclude the achievements, discuss limitations and show how future work can further improve the presented work.

Firsly, the need for evaluating the security of USB key components was identified as being necessary when assessing the security of a USB enabled system in Section 1. The reason being the USB drivers in kernel space, which are a very compelling target for an attacker not only because of the elevated privileges but also because of the wide deployment of USB. Secondly, fuzzing techniques were introduced, classified and discussed in Section 2. Furthermore, it was shown that fuzzing is well suited to expose bugs in kernel level drivers. Thirdly, an overview of USB and its key components was given in Section 3. Three layers were identified as being attackable with the presented framework: USB stack, drivers on top and applications using the exposed functionality. Fourthly, a design to automatically fuzz-test the identified components was presented and work that was necessary to build the architecture was outlined in Section 4. In order to build the framework various free software products had to be patched, including Scapy and QEMU. The former was improved by adding new fields to describe packets used in USB communication. The latter was enhanced with a USB packet filter as well as support for external software USB devices backed by named pipes.

The presented architecture uses a software-based USB device to generate fuzzed packets. As of now, the software device supports enumeration and parts of the mass-storage protocol only. More protocols are needed to penetrate deeper into the drivers.

The architecture is universal in the sense that knowledge of the guest operating system is not required. If Linux is the guest operating system, we do, however, know about the inner workings. Unfortunately, this information is not taken account when building the fuzzed packets.

Finally, obtained results were presented in Section 5 and they identify the design as being capable of finding flaws in USB kernel drivers. Furthermore, USB stacks of different operating systems have been fingerprinted. This knowledge enables attackers to detect which host operating system they are attached to and to subsequently launch targeted attacks.

A limitation of the presented work is the lack of support for the already specified USB-3. While the general framework should also work with USB-3 enabled operating systems, additional features need to be implemented in the virtual USB device in order to test new features of USB-3 such as device initiated communication.

The fully virtualised approach may also not uncover time critical bugs. USB operates in frames of one millisecond and as the fuzzing framework is running in userspace it might not be able to guarantee processing in time. However, the way for easily carrying fuzz-testing with the ability to smartly and precisely fuzz fields of USB packets has been paved and the modular design of the framework makes it easy to be adapted.

References

[ABS07] Arzt-Mergemeier, Joerg; Beiss, Willi; Steffens, Thomas: The digital voting pen at the Hamburg elections 2008: electronic voting closest to conventional voting. In: Proceedings of the 1st international conference on E-voting and identity. Springer-Verlag, Bochum, Germany, pp. 88–98, 2007.

[AI97] Anderson, Don; Inc., MindShare,: USB system architecture. Addison-Wesley Developers Press, Reading Mass., 1997.

[Be05] Bellard, Fabrice: QEMU, a fast and portable dynamic translator. In: Proceedings of the annual conference on USENIX Annual Technical Conference. USENIX Association, Anaheim, CA, pp. 41–41, 2005.

[Bi10a] Biondi, Philippe: , Scapy - a powerful interactive packet manipulation program. https://www.secdev.org/projects/scapy/, August 2010.

[Bi10b] Biondi, Philippe; Raynal, Fred; Martini, Sebastien; Kacherginsky, Peter: Scapy v2.1.1-dev documentation. April 2010.

[Co00] Compaq; Hewlett-Packard; Intel; Lucent; Microsoft; NEC; Philips: , Universal Serial Bus Specification, April 2000.

[CRK05] Corbet, Jonathan; Rubini, Alessandro; Kroah-Hartman, Greg: Linux device drivers. O'Reilly Media, Inc., 2005.

[Hi05] Hirofuchi, Takahiro; Kawai, Eiji; Fujikawa, Kazutoshi; Sunahara, Hideki: USB/IP: a peripheral bus extension for device sharing over IP network. In: Proceedings of the annual conference on USENIX Annual Technical Conference. USENIX Association, Anaheim, CA, pp. 47–60, 2005.

[JJ10] Jodeit, Moritz; Johns, Martin: USB Device Drivers: A Stepping Stone into your Kernel. In: 2010 European Conference on Computer Network Defense. IEEE Computer Society, Berlin, Germany, October 2010.

[KK07] Keil, S.; Kolbitsch, C.: Stateful Fuzzing of Wireless Device Drivers in an Emulated Environment. Black Hat Japan, 2007.

[Le07] Leavitt, Neal: For Wireless USB, the Future Starts Now. Computer, 40(7):14–16, 2007.

[NN08] Neystadt, John; Natanov, Nissim: , Application - Testing Software Applications with Schema-based Fuzzing, April 2008.

[TDM08] Takanen, Ari; DeMott, Jared; Miller, Charles: Fuzzing for Software Security Testing and Quality Assurance. Artech House, 1 edition, June 2008.

[Th06] Thales: , Thales Inflight Entertainment Systems, November 2006.

[Ve08] Venkateswaran, Sreekrishnan: Essential Linux device drivers. Prentice Hall, Upper Saddle River NJ, April 2008.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Mechanismus zur Synchronisation verteilter Datenmodelle für kollaborative Editoren

Sebastian Runge[1]

Abstract: Innerhalb der letzten zehn Jahre haben kollaborative Editoren (Mehrbenutzer-Editoren) an Bedeutung gewonnen. Dies sind Anwendungen, die mehreren Nutzern die gemeinsame Bearbeitung von Daten erlauben, auch wenn die Nutzer zeitlich oder räumlich voneinander getrennt sind. In diesem Beitrag wird ein Mechanismus beschrieben, welcher die Datenmodelle der verschiedenen Nutzer synchronisiert und so die Zusammenarbeit ermöglicht. Am Modell eines Nutzers vorgenommene Änderungen werden über ein Netzwerk unmittelbar an alle anderen Nutzer übertragen. Weiterhin wird es den Nutzern ermöglicht, auch ohne eine Netzwerkverbindung die Daten lokal zu bearbeiten, um sich zu einem späteren Zeitpunkt wieder mit den anderen Nutzern zu synchronisieren.

Keywords: kollaborative Editoren, verteilte Datenmodelle, Modellsynchronisation

1 Einleitung

Innerhalb der letzen zehn Jahre haben kollaborative[2] Editoren, auch als Mehrbenutzer-Editoren bezeichnet, an Bedeutung gewonnen. Dies sind Computerprogramme, die zur sogenannten *Groupware* gehören und in den Bereich des interdisziplinären Forschungsgebiets CSCW (*Computer Supported Cooperative Work*) fallen. Kollaborative Editoren erlauben es ihren Nutzern, gemeinsam Dokumente wie Texte oder Bilder zu bearbeiten, auch wenn die Nutzer zeitlich oder räumlich voneinander getrennt sind. Diese Zusammenarbeit kann entweder zeitversetzt wie bei einem Versionsverwaltungssystem erfolgen oder mit einer Synchronisation in Echtzeit, bei der das Editieren eines Nutzers sofort auch auf den Bildschirmen der anderen Nutzer sichtbar ist. Ermöglicht wird diese Technik durch die starke Vernetzung moderner Computersysteme. Erforderlich wird sie durch die Tatsache, dass heutige Produktionsprozesse, nicht nur die Softwareentwicklung, häufig in einem Team stattfinden.

Ein geläufiges Anwendungsgebiet für kollaborative Editoren ist das gemeinsame Schreiben von Texten, wie es etwa in Google Docs[3] oder Etherpad[4] ermöglicht wird. Diese Anwendungen nutzen eine Synchronisationstechnik namens operationale Transformation (OT) [EG89], wobei die von Neil Fraser in [Fr09] vorgestellte differentielle Synchronisation eine Alternative dazu darstellt. Die genannten Techniken sind jedoch in erster Linie

[1] ehemals Universität Kassel, egnurs@gmail.com

[2] Im Deutschen wird unter dem Wort Kollaboration eigentlich die Zusammenarbeit mit dem Feind zu Kriegszeiten verstanden. In diesem Beitrag bezieht sich der Begriff jedoch auf die Zusammenarbeit im Allgemeinen, angelehnt an das englische und französische *collaboration*.

[3] https://www.google.de/intl/de/docs/about/

[4] http://etherpad.org/

auf die Synchronisation von Texten ausgelegt, während in diesem Beitrag die Synchronisation von allgemeinen Modelldaten behandelt werden soll.

Die Elemente, welche die Nutzer über die Benutzeroberfläche (GUI) eines Editors bearbeiten, werden durch ein Datenmodell im Arbeitsspeicher repräsentiert. Die Modelldaten bestehen, zumindest bei Anwendung des objektorientierten Programmierparadigmas, aus Objekten mit Attributen sowie aus Referenzen zwischen den Objekten. Im Falle des kollaborativen Editierens, bei dem die Nutzer in der Regel an verschiedenen Computern arbeiten und gleichzeitig ein Dokument editieren, verfügt jeder Nutzer über eine eigene Instanz beziehungsweise eine Kopie dieses Objektmodells. Damit die Zusammenarbeit der Nutzer funktioniert, müssen Änderungen, die ein Nutzer an seiner Modellinstanz vornimmt, auch an den Modellinstanzen aller anderen beteiligten Nutzer wirksam werden. Dazu ist ein Mechanismus notwending, der die Modellinstanzen mit Hilfe einer Netzwerkverbindung untereinander synchronisiert. Ein solcher Mechanismus, der im Rahmen einer Masterarbeit entstanden ist, wird in diesem Beitrag vorgestellt.

Ein Nachteil der operationalen Transformation ist, dass sie als nicht einfach gilt. Joseph Gentle, ein ehemaliger Entwickler der OT-Anwendung Apache Wave schreibt etwa: „Die korrekte Implementierung der Algorithmen ist sehr schwierig und zeitaufwändig“ [Ge11]. Die differentielle Synchronisation hingegen hat den Nachteil, dass mehrere Kopien eines Datenmodells im Speicher gehalten werden müssen, damit die Differenzen zwischen diesen ermittelt werden können. Daher wurde in der Masterarbeit auf beide Ansätze verzichtet mit dem Ziel, den Synchronisationsmechanismus möglichst simpel zu gestalten. Die Synchronisation soll dabei unsichtbar für alle Anwender durchgeführt werden. Auftretende Merge-Konflikte sollen automatisiert aufgelöst werden, sodass der Editierfluss der Nutzer nicht durch Konfliktmeldungen und manuelles Merging gebremst wird. Auch die Entwickler der kollaborativen Editoren sollen sich nicht mit der Synchronisation auseinandersetzen müssen: Die Synchronisation des gesamten Modells soll ohne weiteren Programmieraufwand geschehen, sobald das Root-Objekt des Modells in den Mechanismus eingehängt wurde.

Der Mechanismus soll zwei Betriebsmodi unterscheiden können: den Online- sowie den Offline-Modus. Im Online-Modus arbeiten mehrere Nutzer zeitgleich, wobei Modelländerungen direkt über eine Netzwerkverbindung übertragen werden. Im Offline-Modus besteht keine Verbindung zu anderen Nutzern; ein Nutzer kann dann trotzdem die Daten weiter bearbeiten und sich zu einem späteren Zeitpunkt wieder mit den anderen Nutzern synchronisieren.

Im weiteren Verlauf werden Aufbau und Funktionsweise des Synchronisationsmechanismus behandelt (Abschnitt 2). Im Anschluss wird als Beispielanwendung ein kollaborativer UML-Editor vorgestellt (Abschnitt 3). Abschließend folgen das Fazit sowie der Ausblick auf mögliche Verbesserungen und Erweiterungen (Abschnitt 4).

2 Der Synchronisationsmechanismus

Die Implementierung des Synchronisationsmechanismus erfolgt in der Programmiersprache Java unter Zuhilfenahme der Bibliothek SDMLib[5]. SDMLib ist eine im Fachgebiet Software Engineering der Universität Kassel entwickelte Bibliothek zum Generieren von Klassenmodell-Code in Java. Weiterhin bietet SDMLib eine Objektverwaltung, bei der Objekten global eindeutige Identifikationsnummern (Objekt-IDs) zugewiesen werden sowie Änderungen an diesen Objekten erfasst und protokolliert werden können. Das grundlegende Verfahren bei der Objektverwaltung wird in [Li12] beschrieben.

2.1 Grundlegender Aufbau

Abb. 1 zeigt den grundlegenden Aufgbau des Mechanismus, der im Folgenden erläutert wird. Gezeigt werden die beiden Nutzer Alice und Bob, jedoch funktioniert der Mechanismus mit beliebig vielen Nutzern.

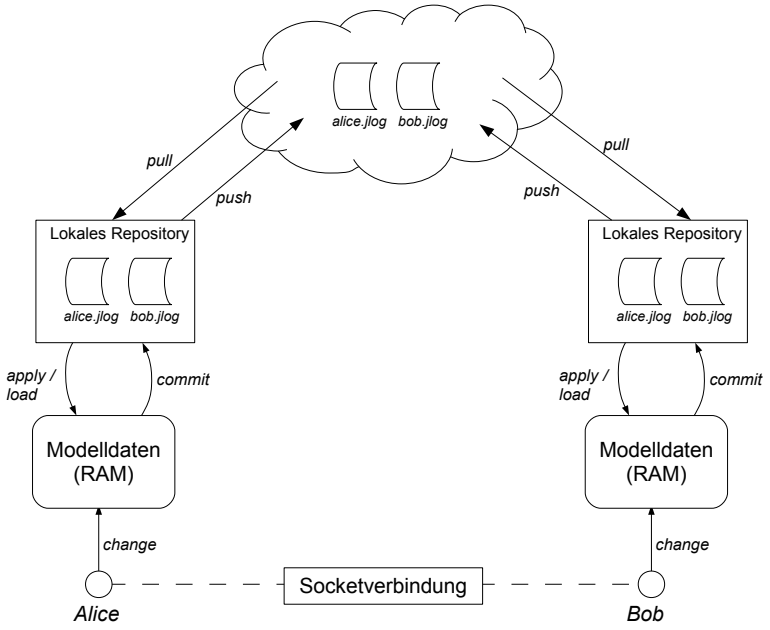


Abb. 1: Grundlegender Aufbau des Synchronisationsmechanismus

Die Synchronisation der Modelldaten geschieht ohne Zutun des Benutzers und wird in regelmäßigen Abständen automatisch wiederholt. Die Architektur ist darauf ausgelegt, dass es keinen zentralen Server gibt, der die Synchronisation vornimmt; stattdessen sollen

[5] https://github.com/fujaba/SDMLib

sich die Nutzer zu einem Peer-to-Peer-Netzwerk zusammenschließen und auch arbeiten können, wenn keine Netzwerkverbindung verfügbar ist. Die folgenden Punkte erläutern die einzelnen Komponenten aus Abb. 1 genauer:

- *Modelldaten (RAM):* Jeder Nutzer, der gerade eine Datenbasis, etwa ein Dokument, im kollaborativen Editor bearbeitet, verfügt über eine Kopie der Modelldaten in seinem Arbeitsspeicher (RAM). Änderungen, die ein Nutzer in der Editor-GUI vornimmt, werden am Modell wirksam.

- *Lokales Repository:* Jeder Nutzer verfügt über ein lokales Repository. Dabei handelt es sich letztlich um einen Ordner auf der eigenen Festplatte, in dem sich Change-Log-Dateien befinden. Diese Log-Dateien enthalten die *History* (Geschichte) der Änderungen, welche die Nutzer an den Modelldaten bislang vorgenommen haben. Die Dateien haben die Endung *.jlog*, was für „JSON-Log“ steht und ein eigenes Dateiformat ist. Der Vorteil des lokalen Repositorys ist, dass jeder Nutzer über den Datenbestand verfügt. Somit können alle Nutzer unabhängig voneinander offline arbeiten. Das Repository enthält eine indirekte Persistierung des Datenmodells, da durch das Nachvollziehen aller Änderungen das gesamte Modell rekonstruiert werden kann. Siehe dazu auch die Operation *Load* weiter unten.

- *Globales Repository:* Zusätzlich zu den lokalen Repositorys der einzelnen Nutzer gibt es ein globales (zentrales) Repository, auf das alle Nutzer Zugriff haben. In Abb. 1 ist es durch eine Wolke (*Cloud*) symbolisiert. Das globale Repository kann ein Netzwerkordner sein, auf den jeder Benutzer Zugriff hat, oder auch ein Cloud-Storage von Dienstleistern wie Dropbox[6] oder Owncloud[7]. Als effektives Mittel für die Synchronisation von Ornderinhalten hat sich auch BitTorrent Sync[8] erwiesen. Bei all diesen Alternativen ist der Administrationsaufwand verhältnismäßig gering, da kein Server eingerichtet werden muss; stattdessen muss lediglich ein Ordner auf der Festplatte eines Nutzers freigegeben oder ein Cloud-Storage-Client installiert werden. Das globale Repository enthält ebenfalls die jlog-Dateien. Über das globale Repository werden die lokalen Repositorys miteinander synchronisiert, das heißt es wird dafür gesorgt, dass alle lokalen Repositorys möglichst immer die gleichen Dateien enthalten. Um Zugriffskonflikte auf die Dateien im globalen Repository auszuschließen, wird sichergestellt, dass jeder Nutzer seine Änderungen in eine eigene jlog-Datei schreibt, die nach dem Nutzer benannt ist. Entsprechend schreibt Alice nur in die Datei *alice.jlog* und Bob nur in *bob.jlog*. Dadurch sind die einzelnen Log-Dateien disjunkt zueinander, das heißt eine konkrete Änderung am Datenmodell tritt nur in einer einzigen der Dateien auf. Eine Datei enthält somit immer nur einen Teil der History; die vollständige History beziehungsweise die vollständigen Modellinformationen ergeben sich nur, wenn die Dateien aller Nutzer vorliegen.

[6] http://www.dropbox.com/
[7] http://owncloud.org/
[8] http://www.getsync.com/

Zwischen diesen genannten Komponenten werden Operationen durchgeführt, die in Abb. 1 durch beschriftete Pfeile dargestellt werden. Die folgenden Auflistung erklärt diese Operationen genauer:

- *Commit:* Die Übertragung einer einzelnen Änderung vom Arbeitsspeicher in das lokale Repository. Ein *Commit* wird unmittelbar nach Auftreten einer Änderung durchgeführt: Unter Verwendung des Observer-Patterns ([St11], S. 169) sind an sämtlichen Attributen im Objektmodell Listener angemeldet, die jede Änderung registrieren, sobald sie am Modell vorgenommen wird. Eine Modelländerung kann dabei das Erzeugen eines Objekts darstellen oder auch die Änderung eines Attributwerts eines bereits existierenden Objekts. Aus jeder Änderung wird ein Change-Objekt erzeugt, welches alle relevanten Informationen der Änderung enthält. Bei einem *Commit* wird ein Change-Objekt in das textuelle JSON-Format persistiert und in die jlog-Datei des Nutzers im lokalen Repository des Nutzers geschrieben.
- *Apply:* Die Übertragung einer einzelnen Änderung vom lokalen Repository in den Arbeitsspeicher. Ein *Apply* ist die Anwendung einer Änderung eines anderen Nutzers am eigenen Datenmodell. Dazu werden die jlog-Dateien der anderen Nutzer im lokalen Repository regelmäßig ausgelesen. Wird dabei eine Änderung gelesen, die in der eigenen History noch nicht bekannt ist, wird diese Änderung am eigenen Datenmodell angewandt und der eigenen History hinzugefügt, nicht jedoch in die eigene jlog-Datei eingetragen. Wenn die Änderung am eigenen Datenmodell angewandt wird, würde sie jedoch wie eine eigene Änderung von einem Listener registriert werden und einen *Commit* verursachen. Da jedoch nur eigene Änderungen in die eigene jlog-Datei eingetragen werden sollen, wird vor dem *Apply* einer fremden Änderung die *Commit*-Funktionalität mit einem einfachen Boolean-Flag vorübergehend außer Kraft gesetzt.
- *Load:* Das Wiederherstellen des Datenmodells im Arbeitsspeicher anhand der Log-Dateien im lokalen Repository. Um die Modelldaten wieder in den Arbeitsspeicher zu laden, zum Beispiel nach einem Neustart des Editors, werden alle zur Datenbasis gehörenden jlog-Dateien im lokalen Repository eingelesen, wobei alle Änderungen in eine gemeinsame Liste geschrieben werden. Die Liste wird anschließend nach logischen Zeitstempeln der Änderungen sortiert und die Änderungen werden in dieser Reihenfolge am Modell angewandt, sodass das Modell Schritt für Schritt rekonstruiert wird. Das Wiederherstellen des Datenmodells, das Zusammenfügen der jlog-Dateien sowie die Behandlung von dabei auftretenden Merge-Konflikten wird in Abschnitt 2.2.1 genauer beschrieben.
- *Push:* Die Übertragung der eigenen jlog-Datei vom lokalen in das globale Repository. Durch einen *Push* werden die eigenen Änderungen den anderen Nutzern bekannt gemacht. Dazu wird die eigene jlog-Datei vom lokalen Repository in das globale Repository kopiert, insofern das globale Repository erreichbar ist. So würde Nutzerin Alice etwa die Datei *alice.jlog* vom lokalen in das globale Repository kopieren; ist im globalen Repository bereits eine Datei *alice.jlog* vorhanden, so wird diese überschrieben.

- *Pull:* Die Übertragung der fremden jlog-Dateien vom globalen Repository in das eigene lokale Repository. Durch einen *Pull* werden die Änderungen der anderen Nutzer abgeholt. Dazu werden die jlog-Dateien - mit Ausnahme der Datei, die dem Nutzer zugeordnet ist - aus dem globalen Repository in das lokale Repository kopiert, insofern das globale Repository erreichbar ist. Nutzerin Alice etwa würde die Datei *bob.jlog* vom globalen in das lokale Repository kopieren; ist im lokalen Repository von Alice bereits eine Datei *bob.jlog* vorhanden, so wird diese überschrieben. Direkt nach dem *Pull* werden die fremden jlog-Dateien im aktualisierten lokalen Repository ausgelesen, um nach noch nicht bekannten Änderungen der anderen Nutzer zu suchen. Diese werden ähnlich zu *Load* gesammelt, sortiert und in zeitlich richtiger Reihenfolge am eigenen Modell angewandt (*Apply*). Die eigene jlog-Datei im lokalen Repository muss an dieser Stelle selbstverständlich nie ausgelesen werden, da sie nur Änderungen enthält, die bereits bekannt beziehungsweise an der eigenen Modellinstanz umgesetzt sind.

Diese Operationen werden nie vom Benutzer bewusst ausgelöst, sondern finden automatisch im Hintergrund statt. Während die Operation *Commit* immer dadurch ausgelöst wird, dass ein Nutzer eine Änderung vornimmt, werden die Operationen *Push* und *Pull* in einem festen Zeitintervall wiederholt durchgeführt; vorstellbar ist eine Durchführung in jeder Minute oder in einem kürzeren Zeitraum. *Push* und *Pull* dienen zur Synchronisation des globalen Repositorys mit den lokalen Repositorys beziehungsweise letztlich zur Synchronisation der lokalen Repositorys untereinander. Da jeder Nutzer nur die eigene jlog-Datei verändert, ist zudem gesichtert, dass *Push* und *Pull* jederzeit unabhängig voneinander und in beliebiger Reihenfolge durchgeführt werden können.

Bei den angegebenen *push/pull*-Intervallen von über einer halben Minute ist eine Echtzeit-Synchronisation natürlich nicht möglich. Daher gibt es die in Abb. 1 gezeigte Komponente *Socketverbindung*. Dabei handelt es sich um eine direkte Netzwerkverbindung zwischen den Nutzern, die eingerichtet wird, sobald erkannt wird, dass die Nutzer zeitgleich mit dem Editor arbeiten. Diese Verbindung dient zur Online-Synchronisation zwischen den Nutzern, die deutlich schneller und unmittelbarer ist als die die Offline-Synchronisation über die Repositorys. Dieser Online-Modus wird in Abschnitt 2.3 genauer erläutert.

2.2 Die Change-Log-Dateien

Listing 1 zeigt beispielhaft eine einzelne Änderung im JSON-Format, wie sie in der Praxis in einer jlog-Datei zu finden wäre. Der JSON-Code wird ohne die Zeilenumbrüche in die Datei geschrieben, sodass jedes Change-Objekt genau eine Zeile in der Datei einnimmt.

```
{
  "class" : "org.sdmlib.replication.ReplicationChange",
  "id"    : "alice.R51",
  "prop"  :
  {
    "historyIdPrefix" : "alice",
```

```
      "historyNumber"   : 23,
      "targetObjectId"  : "bob.S27",
      "targetProperty"  : "name",
      "changeMsg"       :
      {
        "id"  : "bob.S27",
        "upd" :
        {
          "name" : "Otto"
        }
      }
    }
}
```

List. 1: jlog-Eintrag einer einzelnen Änderung am Modell

In der gezeigten Änderung hat Nutzerin Alice den Wert eines Namensattributs mit „Otto“ überschrieben. Die ID der Änderung, *alice.R51*, ist zusammengesetzt aus dem *historyId-Prefix* (hier *alice*) und einer fortlaufenden Nummer (hier *51*). Die fortlaufende Nummer gibt die zeitliche Abfolge der Änderungen vor. Der Präfix gibt nicht nur den Urheber der Änderung an, sondern stellt sicher, dass die Änderungs-IDs systemübergreifend einmalig sind: Wenn die Nutzer Alice und Bob ohne Verbindung zueinander arbeiten, können sie Änderungen mit gleicher fortlaufender Nummer erzeugen - dies kann in selteneren Fällen auch im Online-Modus passieren. Durch das Hinzufügen des Nutzernamens zur ID können jedoch die Änderungen *alice.R51* und *bob.R51* voneinander unterschieden werden.

Die *targetObjectId* (hier *bob.S27*) entspricht der ID des Objektes, auf welches sich die Änderung bezieht. Wie zu sehen ist, sind die IDs von Objekten im Modell genau wie die Änderungs-IDs aufgebaut; der Präfix gibt an, dass das Objekt ursprünglich von Nutzer Bob erzeugt worden ist. Die *targetProperty* schließlich gibt den Namen des Attributs an, das verändert wird.

2.2.1 Wiederherstellung des Datenmodells

Zur Rekonstruktion des Objektmodells müssen die Änderungslisten der Nutzer gemischt (*merge*) werden. Dazu werden die Änderungen aus den einzelnen jlog-Dateien ausgelesen und in einer gemeinsamen Liste gespeichert. Die Änderungen in der Liste werden nun nach ihrer zeitlichen Abfolge, also nach der fortlaufenden Nummer in der Objekt-ID, sortiert. Sollten zwei Nummern gleich sein, so wird nach den Nutzernamen alphabetisch sortiert. Abb. 2 zeigt ein Merge-Beispiel mit den Änderungen der Nutzer Alice und Bob, wo die Änderung *alice.R24* zeitlich vor der Änderung *bob.R24* sortiert wird, da Alice weiter vorne im Alphabet als Bob liegt.

Sobald die Liste sortiert ist, können die Änderungen in dieser Reihenfolge angewandt werden. Haben mehrere Nutzer das gleiche Attribut eines Objekts verändert, liegt ein Merge-

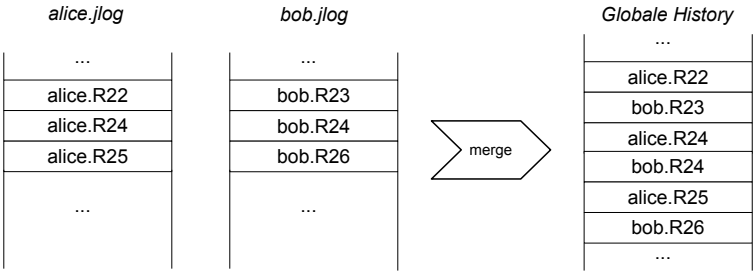


Abb. 2: Einfaches Merge-Beispiel mit ID-Nummern-Überschneidung

Konflikt vor. Ein solcher wird durch die Sortierung jedoch automatisch aufgelöst: Enthält die Änderungsliste mehrere Änderungen am gleichen Objekt und Attribut, so führt die konsekutive Anwendung der zeitlich sortierten Änderungen dazu, dass der Wert jenes Attributs mehrfach überschrieben wird und die zeitlich letzte Änderung an diesem Attribut schließlich den finalen Wert bestimmt.

2.3 Der Online-Modus

Sobald mehrere Nutzer zeitgleich eine Datenbasis bearbeiten, wird eine Socket-Verbindung zwischen ihnen aufgebaut und der Online-Modus gestartet. Dies ermöglicht eine unmittelbare Zusammenarbeit zwischen den Nutzern durch eine direkte Synchronisation der Modelldaten sowie der Übertragung zusätzlicher Nutzerdaten wie die Mauscursor-Positionen. Die Übertragung zusätzlicher Informationen ist dabei vom Synchronisationsmechanismus entkoppelt, sodass etwa die Mauscursor-Positionen nicht in die History aufgenommen werden und so mit höheren Update-Raten als andere Änderungen übertragen werden können. Damit alle Teilnehmer des Online-Modus von Beginn an synchron sind, tauschen die Editoranwendungen direkt nach dem Einrichten der Verbindung ihre Änderungshistorien untereinander aus.

2.3.1 Verbindungsherstellung

Auch wenn die Socket-Verbindung besteht, wird weiterhin die Synchronisation über die Repositorys sowie das regelmäßige Auslesen der jlog-Dateien durchgeführt. Zwar sind die Änderungen, die auf diese Weise registriert werden, in der Regel bereits durch die schnellere Online-Synchronisation bekannt und können verworfen werden; jedoch können die Repositorys zur Verbindungsherstellung genutzt werden, wenn Nutzerinformationen Teil des synchronisierten Datenmodells sind. So kann zum Beispiel für jeden Nutzer eine Boolean-Variable im Modell angelegt werden, die angibt, ob ein Nutzer gerade online (*true*) oder

offline (*false*) ist, das heißt ob ein Nutzer gerade den Editor verwendet sowie Netzwerkzugriff auf das globale Repository hat. Dazu kommen Attribute für das Hinterlegen von IP-Adresse und Port-Nummer. Sobald ein Nutzer über die Repositorys registriert, dass jenes *Online-Flag* eines anderen Nutzers auf *true* gesprungen ist, kann mit den Angaben von IP-Adresse und Port eine Socket-Verbindung zu diesem Nutzer hergestellt werden. Der Übergang vom Offline-Modus in den Online-Modus oder umgekehrt geschieht dabei fließend und im laufenden Betrieb.

2.3.2 Behandlung von Übertragungsfehlern

Bei der Übertragung der Änderungen über das Netzwerk kann es zu Fehlern kommen, sodass zum Beispiel Nachrichten verloren gehen und nie empfangen werden. Damit die Datenmodelle der verschiedenen Nutzer jedoch synchron zueinander bleiben, ist es nötig, dass jede einzelne Änderung von jedem Nutzer erfasst wird. Für die Echtzeit-Synchronisation im Online-Modus werden verlorengegangene Änderungen daher erneut über das Netzwerk angefordert.

Dazu enthält eine Change-Nachricht, die über das Netzwerk verschickt wird, eine Referenz auf die vorhergehende Nachricht. Wird eine Change-Nachricht empfangen, wird aus ihr die ID der vorhergehenden Nachricht ausgelesen. Ist diese ID in der eigenen History nicht verzeichnet, bedeutet dies, dass mindestens eine Nachricht verpasst wurde. In diesem Fall wird das nochmalige Senden der vorhergehenden Änderung angefordert. Dieser Vorgang wird rekursiv wiederholt, bis alle verpassten Nachrichten wieder eingefangen wurden. Diese Behandlung von Übertragungsfehlern wird dabei nur bei der Übertragung von Modelländerungen, nicht aber bei der Übertragung von GUI-Informationen wie Mauscursor-Positionen angewandt.

2.3.3 Netzwerktopologie

Die in Abschnitt 2.3.1 beschriebene Methode zur Verbindungsherstellung führt dazu, dass jeder Nutzer eine Netzwerkverbindung zu jedem anderen Nutzer aufbaut. Dies entspricht der Netzwerktopologie des vollständigen Graphen (*Full Mesh*). Diese Topologie hat zunächst den Vorteil der Adaptivität: Bei den Verbindungen beim kollaborativen Editieren handelt es sich schließlich nicht um ein Netzwerk mit konstanter Teilnehmerzahl, denn es können sich jederzeit Nutzer abmelden oder anmelden. Beim Full Mesh ist es recht einfach, im laufenden Betrieb einen neuen Teilnehmer in das Netz zu integrieren oder einen bestehen Teilnehmer zu entfernen, da dazu nur Verbindungen eingerichtet oder entfernt werden müssen, die nur jenen Teilnehmer betreffen. Bei einer Ring-Topologie hingegen müsste eine zusätzliche Softwarekomponente sicherstellen, dass beim An- und Abmelden von Nutzern die Ringstruktur bestehen bleibt.

Ein weiterer Vorteil der Full-Mesh-Topologie ist die hohe Ausfallsicherheit, denn der vollständige Graph hat den höchsten Zusammenhangsgrad. Der Zusammenhangsgrad gibt an, wieviele Leitungen unterbrochen werden können, ohne dass ein Teilnehmer von der

Kommunikation im Netz ausgeschlossen wird ([Ka96], S. 263ff.). Wenn eine direkte Verbindung zu einem Nutzer ausfällt, können diesen auch weiterhin Nachrichten über Umwege erreichen. Zwar gibt es hierfür kein Routing, doch werden Änderungsnachrichten ähnlich zum sogenannten *Flooding* ([Ta03], S. 393f.) im Netzwerk verteilt, indem noch nicht bekannte empfangene Nachrichten auch an die benachbarten Knoten weiterversendet werden. Zu beachten ist hier, dass sich das Flooding negativ auf die Performance auswirken kann und bei hohen Nutzerzahlen möglicherweise eingestellt werden muss.

3 Anwendungsbeispiel: UML-Editor

Um die praktische Anwendbarkeit des Synchronisationsmechanismus zu überprüfen, wurde eine Reihe von Beispielanwendungen in JavaFX entwickelt. Eine davon ist ein kollaborativer UML-Editor, der hier kurz vorgestellt werden soll. Abb. 3 zeigt einen Screenshot des Editors.

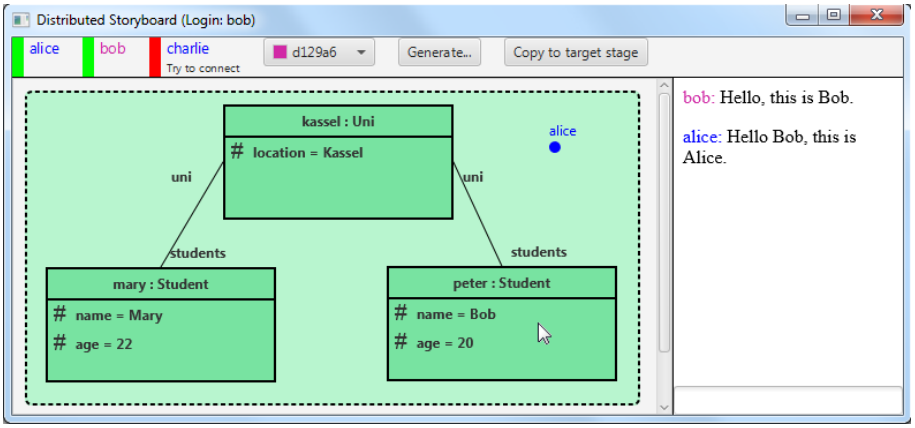


Abb. 3: Screenshot des UML-Editors

Mit dem UML-Editor kann ein Objektdiagramm gezeichnet werden, indem Objektkästen erzeugt werden, die mit Attributen sowie Attributwerten versehen werden können und mit Referenzen untereinander verknüpft werden können. Aus diesem Objektdiagramm wird dann ein Klassenmodell abgeleitet, für welches auch Java-Code generiert werden kann. Im Rahmen dieses Beitrags interessant sind jedoch hauptsächlich die kollaborativen Elemente des Editors.

In einer Leiste am oberen Fensterrand sind alle Nutzer aufgelistet, die an der Bearbeitung des UML-Diagramms beteiligt sind. In Abb. 3 sind dies die Nutzer Alice, Bob und Charlie. Ein farbiger Balken neben dem Nutzernamen gibt an, ob der Nutzer derzeit online (grün) oder offline (rot) ist. Hier sind Alice und Bob jeweils online und arbeiten daher im Online-Modus zusammen, während Charlie gerade offline ist. Gehen Nutzer online oder offline, ändert sich die Farbe der Balken zur Laufzeit entsprechend, sodass allen Nutzern stets bewusst ist, mit welchen anderen Nutzern sie derzeit zusammenarbeiten.

Abb. 3 zeigt den UML-Editor aus der Sicht von Nutzer Bob. Sein Mauszeiger ist der des jeweiligen Betriebssystems (hier Windows 7). Zusätzlich werden auch die Mauszeiger der anderen Nutzer in der GUI als mit dem Nutzernamen beschriftete farbige Punkte dargestellt: In Abb. 3 ist als Beispiel Alices Mauszeiger zu sehen. Durch die Darstellung der anderen Mauszeiger ist es einem Nutzer ersichtlich, welche Elemente derzeit von den anderen Nutzern bearbeitet werden. Auch können Nutzer so auf bestimmte Elemente zeigen, während sie zum Beispiel über IP-Telefonie miteinander kommunizieren.

Als weiteres Kommunikationsmittel zwischen den Nutzern dient eine einfache Chatleiste auf der rechten Seite des Fensters. Über ein Color-Picker-Element an der oberen Fensterleiste kann jeder Nutzer eine Farbe für sich wählen und jederzeit ändern. In dieser Farbe werden der Nutzername in der Nutzerleiste, der Nutzername in den Chatnachrichten sowie der Mauszeiger des Nutzers dargestellt. Die Idee hinter dem Farbschema ist, dass jeder Nutzer so Handlungen anderer Nutzer schneller ihrem Urheber zuordnen kann. So sind durch die unterschiedlichen Farben etwa die Mauszeiger schneller voneinander unterscheidbar.

Weiterhin wird das UML-Diagramm zwischen den Nutzern synchronisiert, bei den Nutzern im Online-Modus auch in Echtzeit: Wenn Alice etwa einen neuen Objektkasten in der Editoroberfläche platziert, ein neues Attribut hinzufügt oder eine neue Kante zieht, sind diese Änderungen auch sofort in Bobs GUI sichtbar. Weiterhin werden bidirektionale JavaFX-Bindings [Ja15] genutzt, um die Textfelder im Editor an die entsprechenden String-Attribute im Datenmodell zu binden, sodass der Inhalt dieser Felder buchstabenweise synchronisiert wird. Wenn also zum Beispiel Bob einen Rollennamen einer Referenz ändert, ist der Tippvorgang an dieser Stelle auch in Alices GUI sichtbar.

Für eine effektive Zusammenarbeit ist es wichtig, dass das Layout des UML-Diagramms in den GUIs aller Nutzer gleich ist. Daher wurden die x- und y-Position der Objektkästen ebenfalls mit in das Datenmodell aufgenommen, sodass auch Änderungen an diesen Positionsdaten durch den Synchronisationsmechanismus an alle Nutzer geschickt werden. Dies bedeutet, dass wenn etwa Alice einen Objektkasten des Diagramms auf ihrem Bildschirm mit der Maus verschiebt, diese Verschiebungsbewegung auch in der GUI von Bob zu sehen ist. Die Linien, welche die Referenzen darstellen, sind an den Kästen verankert und verschieben sich dadurch in allen GUI-Instanzen automatisch mit.

4 Fazit und Ausblick

Zwar wurden keine systematischen Peformancetests bezüglich Laufzeitverhalten und Objektmenge durchgeführt, jedoch deutet die testweise Nutzung des kollaborativen UML-Editors darauf hin, dass der Synchronisationsmechanismus zumindest für diesen Anwendungsfall gut funktioniert. Das kollaborative Editieren wurde mit drei Personen gleichzeitig getestet, wobei sich die Anwendung nicht langsamer anfühlte als ein Einzelnutzer-Editor. Auch der fliegende Wechsel zwischen Online- und Offline-Modus verhielt sich wie gewünscht.

Für die Zukunft ist eine Reihe von Verbesserungen am Synchronisationsmechanismus denkbar. Eine mögliche Problematik ist, dass die jlog-Dateien immer weiter anwachsen, sodass bei langfristiger Nutzung eines Editors das Auslesen der Dateien zu viel Zeit in Anspruch nehmen kann. Die Dateien enthalten dabei viele alte Änderungen, die für die Rekonstruktion des Datenmodells oft nicht mehr relevant sind. Dementsprechend wäre eine Kompression der Dateien wünschenswert, welche obsolet gewordene Änderungen entfernt und den Umfang der jlog-Dateien entsprechend reduziert.

Eine weitere noch nötige Funktionalität ist das Löschen von Objekten, das in der momentanen Objektverwaltung von SDMLib noch nicht möglich ist. Sobald das Löschen von Objekten aus dem Objektmodell zwischen den Nutzern synchronisiert werden kann, können auch Undo- und Redo-Operationen implementiert werden, die für Editoren in der Regel von wesentlicher Bedeutung sind.

Weiterhin ist ein Problem, dass Änderungen an einem komplexen Datenmodell oft nicht atomar sind, sondern aus mehreren aufeinanderfolgenden Änderungen bestehen. Diese Änderungen sollten zu Gruppen zusammengefasst werden und mit dem Transaktionsprinzip ausgeführt werden können, um die Konsistenz des Datenmodells zu gewährleisten.

Wünschenswert wäre darüberhinaus die Integration von Ressourcen: Die Nutzer sollten in den Editoren Ressourcen wie Bilder, Musikdateien oder ähnliches einfügen können. Dazu muss der Mechanismus derart erweitert werden, dass solche Ressourcen verwaltet und unter den Nutzern verteilt werden können.

Ein wichtiger Punkt ist schließlich noch die Synchronisation sortierter Listen wie etwa Javas `ArrayList`. Die Objektverwaltung von SDMLib unterstützt bislang nur unsortierte Sets. Sortierte Listen sind jedoch wichtig, wenn Objekte zum Beispiel in einer bestimmten Reihenfolge in der GUI dargestellt werden sollen, sodass alle Nutzer eines kollaborativen Editors die Objekte in genau dieser Reihenfolge sehen können.

Literaturverzeichnis

[EG89] Ellis, C. A.; Gibbs, S. J.: Concurrency Control in Groupware Systems. International Conferfence on Management of Data (SIGMOD), 1989.

[Fr09] Fraser, N.: Differential Synchronisation. DocEng 09, 2009.

[Ge11] ShareJS. Live concurrent editing in your app, sharejs.org/, Stand: 25.02.2015.

[Ja15] JavaFX Bindings, doc.oracle.com/javase/8/javafx/api/javafx/beans/binding/Bindings.html, Stand: 25.02.2015.

[Ka96] Kauffels, F.-J.: Lokale Netze. Datacom, Bergheim, 1996.

[Li12] Lindel, S.: Online synchronization of replicated models. CVSM2012, 2012.

[St11] Starke, G.: Effektive Software-Architekturen. Carl Hanser, München, 2011.

[Ta03] Tanenbaum, A. S.: Computernetzwerke. Pearson Studium, München, 2003.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Metadatenextraktion und Vorschlagssysteme im Visual SPARQL Builder

Lukas Eipert[1]

Abstract: Auch wenn der Vorteil semantischer Technologien in der Informatik erkannt wurde, setzen sich diese in anderen Domänen nur zögerlich durch. Um diesen Prozess zu beschleunigen, bedarf es Werkzeugen, die Nicht-Informatikern den Gebrauch des Semantic Web erleichtern. Der Visual SPARQL Builder ist ein neuartiges Werkzeug und ermöglicht es, SPARQL-SELECT-Anfragen graphisch zu modellieren. Damit man mit dem Visual SPARQL Builder arbeiten kann, werden Metadaten über die Struktur der Daten einer semantischen Wissensbasis benötigt. In diesem Artikel wird beschrieben, wie der Visual SPARQL Builder die Strukturinformationen durch SPARQL-Anfragen ermittelt und dazu benutzt, durch geeignete Vorschlagssysteme die Eingaben für Nutzer zu erleichtern.

Keywords: SPARQL, Semantic Web, Metadaten, Datenexploration, Visuelle Anfragesprache

1 Einleitung

Um Anfragen an semantische Wissensdatenbanken zu stellen, wird häufig SPARQL (SPARQL Protocol and RDF Query Language) genutzt. Die Sprache ist eine W3C-Empfehlung für programmatische Zugriffe auf RDF-Datensätze [Wo13a]. Zahlreiche öffentliche Endpunkte wie die der LOD-Cloud wie die dbpedia bieten Schnittstellen um über SPARQL Daten abzufragen.

Ziel des Projektes ist es, eine Anwendung zu entwickeln, die es ermöglicht, SPARQL-Anfragen einfach graphisch zu modellieren. Durch die Modellierung soll die Komplexität von SPARQL reduziert werden. Diese Anfragen-Modelle werden in SPARQL übersetzt und so kann der Nutzer auf intuitive Weise Anfragen an semantische Wissensdatenbanken stellen. Im Allgemeinen dürfte eine graphische Darstellung komplexer Zusammenhänge in einem diagrammähnlichen Modell leichter zu erfassen und intuitiver sein als ihre textuelle Repräsentation.

Der Visual SPARQL Builder (kurz **VSB**) ist nicht meine Idee, sondern ist aus einem Softwaretechnikpraktikum im Bachelorstudiengang Informatik an der Universität Leipzig hervorgegangen. Das Praktikum wurde von Andreas Nareike und Henri Knochenhauer betreut. Zu diesem Zeitpunkt wurde der VSB noch Graphical SPARQL Builder genannt. Nach dem Ende des Projektes führte ich die als Gruppe angefangene Entwicklung der Anwendung mit dem Ziel weiter, meine Bachelorarbeit über den VSB zu schreiben. Diese ist zur Zeit noch in Bearbeitung und wird 2015 noch fertiggestellt. Derzeit bin ich alleiniger Entwickler des Projektes.

Einige Anwendungsfälle für einen solchen visuellen Editor sind in Abschnitt 2.1 aufgeführt. Aus diesen Anwendungsfällen und den Vorgaben während des Softwaretechnikpraktikums ergeben sich die folgenden Anforderungen an den VSB:

[1] Universität Leipzig, Bachelorstudiengang Informatik, lukas.eipert@studserv.uni-leipzig.de

1. Clientseitiger Editor, welcher im Webbrowser des Anwenders läuft
2. Extraktion der benötigten Metadaten vom genutzten SPARQL-Endpunkt während der Laufzeit der Anwendung
3. Intuitive Benutzeroberfläche mit einem geeigneten Vorschlagssystem, um die Erstellung von Anfragen zu vereinfachen

Um diese Ziele zu erreichen, wurde für den VSB eine diagrammbasierte Modellierungssprache ähnlich visueller Programmiersprachen entwickelt. Auf diese wird im Abschnitt 2.2 genauer eingegangen. Die Anwendung ist als Javascript-Webapplikation umgesetzt und kann in jedem modernen Browser genutzt werden. Daraus ergeben sich gewisse Vorteile, der Administrator kann zum Beispiel durch eine Programmaktualisierung des VSB an zentraler Stelle eine neue Version an alle Nutzer verteilen. Genauere Informationen zur technischen Realisierung finden sich in Abschnitt 2.3. Darauf folgt in Abschnitt 2.4 ein kurzer Vergleich mit existierenden SPARQL-Editoren und wie sich der VSB von diesen unterscheidet.

Im Hauptteil des Artikels wird die Metadatenextraktion aus SPARQL-Endpunkten während der Laufzeit einer Anwendung (Abschnitt 3) und das in den VSB eingebaute Vorschlagssystem (Abschnitt 4) beschrieben. Während der Entwicklung des Visual SPARQL Builder wurden und werden zukünftig natürlich weitere Probleme gelöst, welche in meiner noch fertigzustellenden Bachelorarbeit genauer betrachtet werden.

2 Hintergrund

In diesem Abschnitt werden einige Hintergründe zum Visual SPARQL Builder für ein besseres Verständnis der späteren Abschnitte beleuchtet.

2.1 Anwendungsfälle

Im Wesentlichen gibt es für den VSB zwei Anwendungsfälle. Einerseits die tägliche Anwendung durch Domänenexperten wie Historiker oder Bibliothekare, die aus einem ihnen bereits bekannten Datenbestand schnell Informationen gewinnen möchten. Andererseits für Nutzer, die eine semantische Wissensdatenbank erforschen wollen, jedoch die Struktur und den Datenbestand derselben nicht oder wenig kennen.

In beiden Fällen ist es von Nutzen, dass der VSB weder tiefgreifende SPARQL- noch Vokabularkenntnis erfordert, sondern durch geeignete Metadatenextraktion und Vorschlagssysteme eine intuitive Nutzung semantischer Daten möglich macht.

Der erste Anwendungsfall wird auch im Rahmen meiner Bachelorarbeit näher beleuchtet werden, da der VSB an der Universitätsbibliothek Leipzig benutzt werden soll, um auf Daten des semantischen Electronic Resource Management Systems AMSL[2] zuzugreifen.

[2] http://amsl.technology

Das vereinfachte Erforschen eines Endpunktes kann man anhand der Beispielinstanz des VSB[3] nachvollziehen; diese arbeitet mit dem SPARQL-Endpunkt der dbpedia[4]. Ähnliche Installationen sind auch für jeden anderen SPARQL-Endpunkt denkbar, welcher der Öffentlichkeit Daten zur Verfügung stellt.

2.2 Modellierungssprache

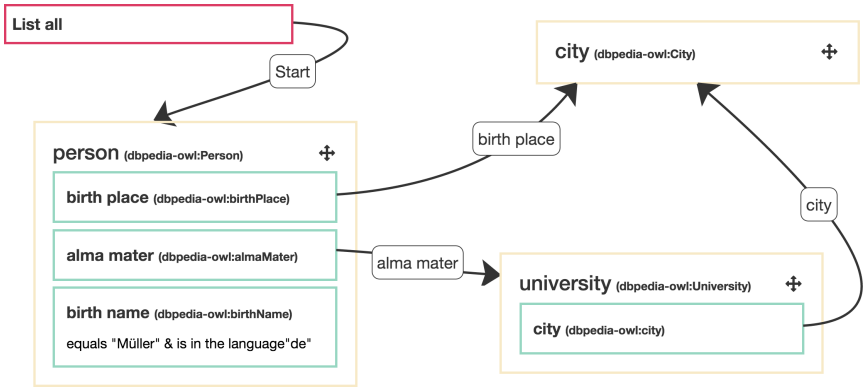


Abb. 1: Modell einer Anfrage an die dbpedia

Wie bereits in der Einleitung erwähnt, wurde für den VSB eine Modellierungssprache ähnlich zur UML [Ob11] entwickelt. In Abb. 1 ist eine Anfrage in dieser Modellierungssprache dargestellt. Es wird beispielhaft nach allen Personen in der dbpedia gesucht, deren Geburtsname “Müller” ist und die in ihrer Geburtsstadt zur Universität gegangen sind. Wie leicht zu erkennen, ähnelt die Sprache einem Datenflussdiagramm.

Die Modellierungssprache wurde im Softwaretechnikpraktikum erstmals entworfen und während des Fortschreitens des Projektes weiterentwickelt. Im Wesentlichen wurden die Anfragesprache SPARQL 1.1 und die Empfehlungen RDF, RDFS und OWL des W3C während der Entwicklung des Sprachmodells berücksichtigt ([Wo14a], [Wo14b], [Wo04]). Daraus entstand diese klassenbasierte Diagrammsprache, in der Klassen durch Eigenschaften eingeschränkt und in Relation gesetzt werden können. Klassen sind als Boxen umgesetzt. Eigenschaften einer Klasse werden in der Klassenbox als Unterboxen dargestellt. Für die Abbildung von Relationen zwischen Klassen werden Pfeile benutzt.

Durch die gewählte Darstellung wird auch die Semantik von Klassen und Eigenschaften besser vermittelt als bei einem puren Knoten-Kanten-Modell. Wenn sowohl Klassen als auch Eigenschaften als Knoten dargestellt werden, ist die Unterscheidung von Relationen zwischen Klassen einerseits sowie Relationen zwischen Eigenschaften und Klassen andererseits nicht so leicht ersichtlich. Wenn man Eigenschaften jedoch als Teil einer Klasse darstellt, ergibt sich dieser Zusammenhang auf einen Blick. Des Weiteren hätte ein pures Knoten-Kanten-Modell viel mehr graphische Elemente als das jetzige Modell, da mehr

[3] http://leipert.github.io/vsb/dbpedia
[4] http://dbpedia.org/sparql

Kanten nötig wären um den Zusammenhang zwischen Eigenschaften und Klassen darzustellen. Es gibt einige visuelle Repräsentationen von SPARQL, die ein Knoten-Kanten-Modell nutzen [vgl. Abschnitt 2.4].

Als besonderes Element der Modellierungssprache ist die Wurzel zu erwähnen. Diese erhöht die Lesbarkeit deutlich, da ein Bezugspunkt für die Betrachtung der Anfrage gegeben ist. Perspektivisch sind auch andere Anfragen als SELECT-Queries denkbar, die dann durch unterschiedliche Wurzelbezeichner ausgedrückt werden können.

Anfragen in der Modellierungssprache des VSB können in JSON (Javascript Object Notation) serialisiert und aus dieser wieder deserialisiert werden. Die Übersetzung nach SPARQL findet aus der serialisierten JSON-Variante eines Anfragenmodells statt.

2.3 Technische Realisierung

Der Visual SPARQL Builder ist als Javascript-Webapplikation umgesetzt. Dabei wurde das von Google entwickelte Javascript-Framework AngularJS[5] verwendet. Zum einen folgt die Entwicklung in AngularJS einem aus dem Studium bekannten Model-View-Controller Paradigma, was die Einstiegshürde für die Entwicklung senkte. Zum anderen lassen sich Komponenten leicht modularisieren, was die Weiterentwicklung, Erweiterung und den Austausch derselbigen vereinfacht. Auch die zahlreich existierenden AngularJS-Komponenten haben die Entwicklung beschleunigt. So greifen zum Beispiel die Übersetzung der Oberfläche, das Routing in der Anwendung oder die Dropdown-Menüs auf Open-Source-Projekte zurück.

Für die Kommunikation mit SPARQL-Endpunkten wird die Javascript-Bibliothek Jassa[6] (**Ja**vascript **S**uite for **S**parql **A**ccess) verwendet, eine Entwicklung von Claus Stadler an der Universität Leipzig im Rahmen des GeoKnow-Projektes [SWL14].

Zahlreiche Teile des VSB sind konfigurierbar gestaltet, sodass nicht unbedingt Javascript- oder AngularJS-Kenntnisse notwendig sind, um den VSB für den Einsatz mit einer eigenen semantische Wissensdatenbank anzupassen.

2.4 Existierende SPARQL-Editoren

Es gibt eine Reihe von unterschiedlichen Ansätzen für SPARQL-Editoren. In der einfachsten Ausführung bietet der Editor nur ein Textfeld zum Eingeben von Anfragen. So bringt zum Beispiel die SPARQL-Engine Virtuoso eine solche Oberfläche mit. Es gibt einige Editoren, welche erweiterte Funktionalitäten wie Syntax-Hervorhebung, Syntaxüberprüfung und Autovervollständigung von Syntax und/oder URIs bieten, so zum Beispiel YASGUI[7].

Die Nutzung eines textuellen Editors setzt natürlich Kenntnisse in SPARQL voraus und ist damit für Anwender ohne diese nur bedingt geeignet.

[5] https://angularjs.org/
[6] https://github.com/GeoKnow/Jassa-Core
[7] http://yasgui.org/

Wie im Modellierungskapitel dargelegt, ist das Ziel des VSB, das Formulieren sinnvoller Anfragen im Kontext einer Wissensbasis zu erleichtern. Dabei liegt der Schwerpunkt vor allem auf einer möglichst aussagekräftigen und einfachen Darstellung. Es ist also nicht das Ziel des VSB, alle möglichen SPARQL-Anfragen darstellen und modellieren zu können. Zur Zeit der Recherche waren keine Editoren bekannt, die dieses Ziel vorrangig verfolgen und den bereits in Abschnitt 1 beschriebenen Anforderungen Genüge leisten.

Es gibt jedoch eine Reihe von graphischen Editoren wie iSPARQL[8] oder Nitelite (mit vSPARQL) [Ru08], die sich zum Ziel gesetzt haben, SPARQL möglichst vollständig abzubilden. So gibt es zum Beispiel die Möglichkeit, direkt zwischen textbasierter Eingabe von SPARQL und iSPARQL zu wechseln. Wenn die graphische Repräsentation einer Anfrage in SPARQL nur die simple zweidimensionale Darstellung derselbigen ist, dann wird die Komplexität der Anfrage für den Ersteller nicht gemindert.

3 Metadatenextraktion

Für den Betrieb des VSB werden verschiedene Informationen über den Aufbau des Datensatzes benötigt. Es gibt verschiedene Möglichkeiten, diese Strukturdaten aus der semantischen Wissensbasis zu gewinnen und bereitzustellen.

3.1 Ansätze der Metadatenextraktion

Im Wesentlichen gibt es drei mögliche Ansätze für die Bereitstellung von semantischen Strukturdaten.

1. Die Struktur wird statisch in Dateien bereitgestellt, zum Beispiel im JSON- oder XML-Format. Das birgt den Nachteil, dass diese statischen Dateien bei Änderungen der Struktur neu generiert werden müssen, um diese Änderungen zu reflektieren. Hat man eine wohldefinierte Struktur, die sich selten ändert, wäre dies die passende Lösung, da die Generierung der statischen Dateien selten stattfindet.
2. Die Struktur wird durch den VSB selbst mit Hilfe von SPARQL-Anfragen an den Endpunkt extrahiert. Dadurch wird gewährleistet, dass der Anwender immer mit dem aktuellen Stand der Metadaten arbeiten kann. Nachteilig wirkt sich dies nur aus, falls komplexe Anfragen sehr lange Berechnungszeiten in Anspruch nehmen und so die Performanz des VSB senken, da der Nutzer auf die Strukturdaten warten muss.
3. Die Vorteile der ersten beiden Arten der Metadatenextraktion ließe sich mit Hilfe eines gesonderten Cache-API kombinieren, welches die Strukturdaten bereitstellt. Dieses könnte in regelmäßigen Abständen komplexe SPARQL-Anfragen im Hintergrund ausführen und die Ergebnisse im JSON- oder XML-Format zur Verfügung stellen. Hierbei ist nachteilig, dass man diese API programmieren und warten muss.

[8] https://github.com/openlink/iSPARQL

Für den VSB wurde der zweite Ansatz gewählt, da dies den geringsten Aufwand für einen Administrator darstellt, eine funktionierende Anwendung bereitzustellen und den Nutzen des VSB für einen vorhandenen SPARQL-Endpunkt zu evaluieren. In der Regel sind die Strukturdaten im Endpunkt vorhanden und lassen sich leicht extrahieren. Die zur Metadatenextraktion dienenden SPARQL-Anfragen können leicht durch eine Konfigurationsdatei angepasst werden. Sollte dieser Ansatz für die Verwendung des VSB nicht die optimale Lösung darstellen, kann der Metadatenextraktionsservice leicht ausgetauscht werden, ermöglicht durch die modulare Programmierung.

3.2 Metadatenextraktion im VSB

Im Folgenden sind die für den erfolgreichen Einsatz des VSB benötigten Metadaten beschrieben. Die Standardimplementierung arbeitet vor allem mit den Schemata RDFS und OWL, welche Empfehlungen vom W3C sind und daher als allgemeine Standards angesehen werden können. Die wichtigsten SPARQL-Queries sind die zur Bestimmung der vorhandenen Klassen und möglichen Eigenschaften einer Klasse.

```
SELECT DISTINCT * WHERE {
  {?uri a rdfs:Class .} UNION {?uri a owl:Class .} .
  OPTIONAL {
    ?uri rdfs:label ?label .
    BIND(LANG(?label) AS ?label_loc)
  }
  OPTIONAL {
    ?uri rdfs:comment ?comment .
    BIND(LANG(?comment) AS ?comment_loc)
  }
  FILTER (
    IF(?comment_loc != "" && ?label_loc != "", ?comment_loc = ?label_loc, true)
  )
}
```

List. 1: Bestimmung der verfügbaren Klassen eines Endpunktes

Die verfügbaren Klassen werden mit Hilfe des Queries in List. 1 bestimmt. Es werden alle Klassen der Typen `rdfs:Class` und `owl:Class` gefunden. Insofern die Klassen Bezeichner `rdfs:label` oder Beschreibungen `rdfs:comment` haben, werden diese zusammen mit ihrer Sprache ausgelesen. Die Bezeichner und Beschreibungen werden für das Vorschlagssystem des VSB benötigt.

In List. 2 ist dargestellt, wie die Eigenschaften einer Klasse ausgelesen werden, im Beispiel die Klasse “Schauspieler” der dbpedia. Zuerst werden alle Überklassen und gleichgestellte Klassen der gewünschten Klasse bestimmt. Dann werden alle Eigenschaften geladen, deren `rdfs:domain` der Klasse, Überklasse oder gleichgestellten Klassen entspricht. Im Beispiel wären dies also nicht nur die Eigenschaften der Klasse “Schauspieler”, sondern auch die der Überklasse “Person”. Dies ist notwendig, da eine Unterklasse immer die Eigenschaften ihrer Überklasse erbt. Für das Vorschlagssystem werden weiterhin die möglichen Zielklassen

```
SELECT DISTINCT
?uri ?range ?type ?label ?label_loc ?comment ?comment_loc
WHERE {
  dbpedia-owl:Actor
  \texttt{rdfs:subClassOf|(owl:equivalentClass|^owl:equivalentClass})*
  ?class .
  ?uri rdfs:domain ?class .
  OPTIONAL { ?uri rdfs:range ?range }
  OPTIONAL { ?uri rdf:type ?type }
  [...]
}
```

List. 2: Bestimmung der Eigenschaften einer Klasse, hier `dbpedia-owl:Actor`

`rdfs:range` und die Art `rdf:type` der Eigenschaften geladen. Wie auch bei den Klassen werden Bezeichner und Beschreibungen der Eigenschaften ausgelesen [vgl. List. 1].

Zusätzlich zu der beschriebenen Methode, Eigenschaften ausgehend von der Klasse zu finden, gibt es eine ähnlich aufgebaute, welche nach inversen Relationseigenschaften sucht. So kann nicht nur Schauspieler $\xrightarrow{\text{Kind}}$ Person, sondern auch Schauspieler $\xleftarrow{\text{Elternteil}}$ Person abgebildet werden. Des Weiteren benutzt der VSB intern Anfragen um Unter- und Überklassen zu bestimmen, sowie mögliche Relationen zwischen zwei beliebigen Klassen. Diese Anfragen sind auch für das Vorschlagssystem von Nöten, welches im Folgenden beschrieben wird.

4 Vorschlagssystem

Die wichtigste Komponente des VSB stellt das Vorschlagssystem dar, da durch dieses fehlerhafte Eingaben vermieden werden können. Auf diese Weise kann die Einstiegshürde zur Nutzung einer semantischen Datenbank gesenkt werden. Die Qualität des Vorschlagssystems hängt von der Qualität der Strukturdefinition und der richtigen Konfiguration der Metadatenextraktion des VSB ab. Sind beide gleichermaßen sichergestellt, ist die Nutzung der folgenden beschriebenen Teilkomponenten des Vorschlagssystems vollständig möglich.

4.1 Klassen- & Eigenschaftssuche

Die Suche nach Klassen und Eigenschaften ist im VSB durch Suchfelder mit Dropdown-Vorschlagsliste realisiert [vgl. Abb. 2]. Wenn man mit der Maus über die i-Symbole in der Vorschlagsliste geht, sieht man zusätzliche Informationen zu den Klassen/Eigenschaften.

Im Hintergrund wird hierfür die Liste der verfügbaren Klassen, beziehungsweise verfügbaren Eigenschaften einer Klasse mit Hilfe der beschriebenen Metadatenextraktion vom Endpunkt in den Arbeitsspeicher geladen. Es werden den Universal Resource Identifiers (URI) der zu durchsuchenden Elemente die Bezeichner und Kommentare in der eingestellten Sprache des VSB zugeordnet. Sind diese nicht in der gewünschten Sprache vorhanden,

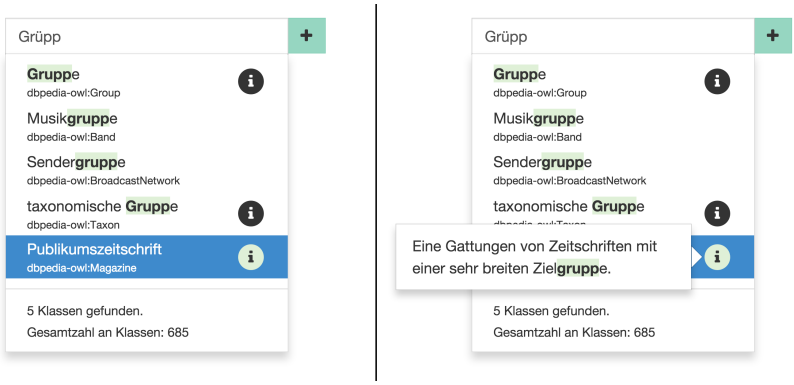


Abb. 2: Beispielhafte Klassensuche mit dem Begriff "Grüpp".

wird eine konfigurierbare Standardsprache wie zum Beispiel Englisch genommen. Ist kein Bezeichner vorhanden, wird einer aus der URI der Ressource ermittelt. So würde der URI “http://dbpedia.org/ontology/Person” der Bezeichner “Person” zugeordnet werden.

Sobald der Nutzer einen Suchbegriff in das Eingabefeld eingibt, wird die Liste der Klassen/Eigenschaften durchsucht und auf alle Elemente reduziert, die einen Treffer in URI, Bezeichner oder Beschreibung aufweisen. Diese werden dann aufsteigend sortiert nach folgenden Kriterien:

1. Der Bezeichner fängt mit dem Suchbegriff an
2. Der Bezeichner enthält den Suchbegriff
3. Restliche Treffer

Jede dieser Gruppen ist in sich alphabetisch nach Bezeichnern sortiert. Die entstandene Liste wird auf eine gewisse Länge gekürzt und dem Nutzer in der Dropdownauswahl ausgegeben, in der die Suchbegriffe hervorgehoben werden.

Mehrere durch Leerzeichen getrennte Suchbegriffe werden als durch den logischen Operator “und” verknüpfte Suchbegriffe angesehen. Alle UTF-8 enkodierten Zeichen werden in der Suche berücksichtigt, bei lateinischen Buchstaben wird jedoch Groß-/Kleinschreibung ignoriert. Buchstaben mit Diakritika wie Akzenten (á), Tremata (ä) oder Tilden (ã) werden durch die Standardvariante (a) ersetzt. Dies soll die Trefferquote bei der Suche erhöhen und Nutzern, die zum Beispiel ein anderssprachiges Tastaturlayout nutzen, die Suche vereinfachen.

In Abb. 2 ist die Arbeitsweise der Suche deutlich zu erkennen. Der Suchbegriff “Grüpp” wurde intern als “grupp” behandelt, und so einige Klassen gefunden. Als erstes ist das Resultat “Gruppe” aufgelistet, da es mit dem Suchbegriff anfängt, gefolgt von einigen Klassen, die den Suchbegriff im Bezeichner enthalten. Als letzten Treffer ist “Publikumszeitschrift” zu finden, da diese Klasse das Wort “Zielgruppe” nur in der Beschreibung enthält.

Die Eigenschaftssuche hat im Unterschied zur Klassensuche die Möglichkeit nach Eigenschaftstypen zu filtern. Außerdem werden die Suchergebnisse nach den Eigenschaftstypen

gruppiert. Um zum Beispiel alle Eigenschaften zu finden, die vom Typ *Zeichenkette* sind, kann der Nutzer einfach ":string" eingeben. Die verfügbaren Kategorien entsprechen den englischen Bezeichnungen der unterstützen Eigenschaftstypen des VSB. Diese werden in Abschnitt 4.2 an den passenden Stellen benannt.

4.2 Eigenschaftsfilter

In SPARQL ist es möglich, Filter auf Werte anzuwenden. Diese Filter unterstützen logische Operatoren, Vergleiche und je nach Datentyp bestimmte Funktionen [Wo13c]. Um jedoch den Nutzern die Eingabe solcher Filter zu vereinfachen und Fehler zu verhindern, bietet der VSB Eingabe- und Auswahlfelder, die semantisch korrekte Eingaben erfordern.

Jeder Eigenschaft im VSB kann der Filter `NOT EXIST` zugewiesen werden, damit man nach allen Instanzen einer Klasse mit Nichtvorhandensein dieser Eigenschaft suchen kann. Des Weiteren kann jede Eigenschaft als optional markiert werden, was kein Filter im Sinne von [Wo13c] darstellt, sondern die komplette Eigenschaft mit Einschränkung gemäß [Wo13b] in einem `OPTIONAL`-Block einfasst.

Der VSB unterscheidet zwischen zwei Eigenschaftstypen. Erstens die Datentypeigenschaften `owl:DatatypeProperty`, welche eine Zielklasse `rdfs:range`) eines elementaren Datentyps wie *Zeichenkette* oder *Zahl* aufweisen. Zweitens die Relationseigenschaften `owl:ObjectProperty`, welche Relationen zwischen Klassen beschreiben. Die Zuordnung von Eigenschaftstypen ist konfigurierbar durch den Administrator des VSB. Zum einen können Eigenschaften nach Typ `rdf:type`, zum anderen nach Zielklasse `rdfs:range` kategorisiert werden. Allen nicht zugeordneten Eigenschaften wird der Typ *Zeichenkette* zugewiesen.

Datentypeigenschaften

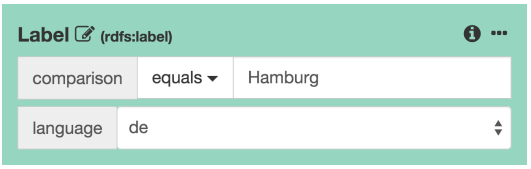


Abb. 3: Zeichenkette mit Filterfeldern

Der VSB unterstützt standardmäßig Datentypeigenschaften der Typen Zeichenkette, Datum und Zahl. Zukünftig sollen auch Uhrzeiten unterstützt werden. Jeder der Datentypen hat eigene Eingabefelder für Filter, die eine Eingabe von syntaktisch falschen Werten verhindern soll:

- **Zeichenketten** haben eine Vergleichssuche mit verschiedenen Optionen wie "enthält" oder "startet mit". Reguläre Ausdrücke sind auch möglich. Zusätzlich lassen sich die Zeichenketten nach ihrer Sprache filtern. In Abb. 3 ist beispielhaft eine String-Eigenschaft mit ausgefülltem Filter dargestellt. Der Kategoriefilter für Zeichenketten ist ":string".

- **Zahlen** können arithmetisch durch Addition, Multiplikation, Subtraktion und Division verändert werden. So ist es ein Leichtes, sich eine Bevölkerungszahl in Tausendern ausgeben zu lassen. Des Weiteren lassen sich die Zahlwerte gemäß der SPARQL-Definition [Wo13c] mit Vergleichsoperatoren wie "<" oder "=" vergleichen. Der Kategoriefilter für Zahlen ist ":number".
- **Daten** lassen sich mit den Operationen "gleich", "ungleich", "vor dem" und "nach dem" vergleichen. Zur Eingabe des Vergleichsdatums wird ein Eingabefeld bereitgestellt, welches das Datum automatisch richtig formatiert und ein Kalender-Widget zur Datumsauswahl bereitstellt. Eine Eingabemöglichkeit für Datenbereiche ist zukünftig ebenso vorgesehen. Der Kategoriefilter für Daten ist ":date".

Datentypeigenschaften können, wenn nötig, von einem Nutzer des VSB in einen anderen Typ überführt werden, damit der passende Filter gesetzt werden kann. Dies ist erforderlich, falls eine Datentypeigenschaft wegen inkorrekter Konfiguration des VSB oder mangelnder Deklaration von `rdfs:range` einer Eigenschaft nicht richtig erkannt wurde. Standardmäßig sind den elementaren Datentypen die entsprechenden Typen der W3C-Empfehlung XSD [Wo12] zugeordnet:

- String: `xsd:string`, `xsd:literal`
- Zahl: `xsd:integer`, `xsd:float`, `xsd:double`, `xsd:decimal`
- Datum: `xsd:date`

Relationseigenschaften

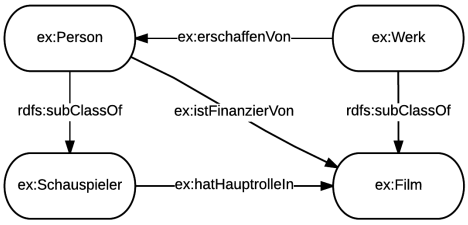


Abb. 4: Graph einer Beispielontologie

Um Relationen zwischen Klassen auszudrücken, sind im VSB Relationseigenschaften implementiert. Es gibt zwei Arten dieser Eigenschaften, die direkten (Kategoriefilter ":object") und die inversen (Kategoriefilter ":inverse"). Anhand von Abb. 4 lässt sich am besten der Unterschied zwischen beiden erklären. In diesem Beispiel kann man der Klasse `ex:Person` eine direkte Eigenschaft `ex:istFinanzierVon` und eine inverse Eigenschaft $\texttt{ex:erschaffenVon}^{-1}$ aus der Beispielontologie "ex:" zuweisen.

Der VSB schränkt die Zuweisung von Relationen auf wohldefinierte mögliche Verbindungen ein. Ist die Beispielontologie aus Abb. 4 im VSB geladen, wäre es nicht möglich die Relation `ex:Schauspieler` $\xrightarrow{\texttt{ex:hatHauptrolleIn}}$ `ex:Werk` zu erstellen, da dies in der Struktur

der Ontologie nicht vorgesehen ist. Wie jedoch in Abschnitt 3.2 bereits erwähnt, beachtet der VSB Klassenhierarchien. So wäre es möglich die Relation `ex:Film` $\xrightarrow{\texttt{ex:erschaffenVon}}$ `ex:Schauspieler`, aber auch ihre Inverse `ex:Schauspieler` $\xleftarrow{\texttt{ex:erschaffenVon}}$ `ex:Film` zu modellieren.

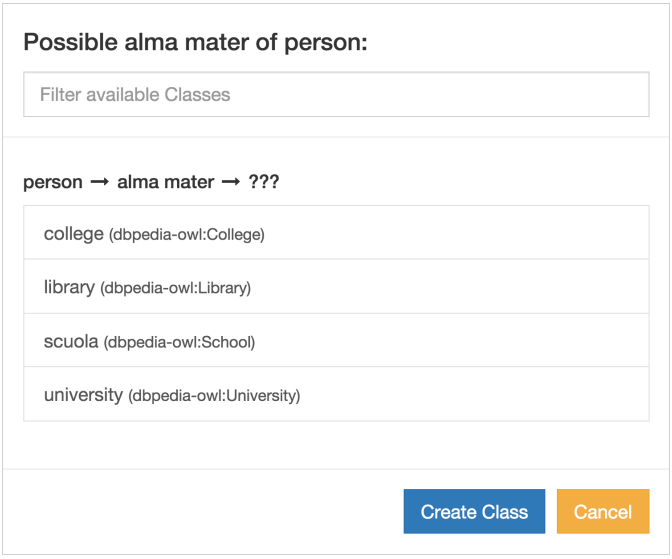


Abb. 5: Zielklassen der Eigenschaft `dbpedia-owl:almaMater`

Um den Nutzern des VSB das Erstellen von Relationseigenschaften zu erleichtern, bietet der VSB zwei Arten von Relationssuchen. Zum einen kann einer existierenden Relationseigenschaft eine passende Klasse hinzugefügt werden, zum anderen kann nach möglichen Relationen zwischen zwei Klassen gesucht werden. In beiden Fällen öffnet sich ein modales Fenster [vgl. Abb. 5], welche dem Nutzer die möglichen Klassen beziehungsweise Eigenschaften vorschlägt. Die Listen in der möglichen Relationen beachtet auch Klassenhierarchien.

5 Fazit und Ausblick

Der VSB wurde derzeit noch nicht in Produktivumgebungen evaluiert. Somit sind auch die Metadatenextraktion als auch das Vorschlagssystem noch nicht ausgiebig getestet. Versuche mit den Ontologien der dbpedia und von AMSL während der VSB-Entwicklung zeigten jedoch vielversprechende Ergebnisse. Im Rahmen meiner Bachelorarbeit wird die Tauglichkeit des VSB für die Ontologie des AMSL-Projektes evaluiert werden. Die Reaktionen auf Vorführungen des Visual SPARQL Builder waren meist positiv.

Das größte während der Entwicklung festgestellte Problem ist die Unverträglichkeit des VSB für mangelnde beziehungsweise implizite Definition der Ontologie. Wenn einer Eigenschaft eine explizite Zuordnung zu einer Klasse durch rdfs:domain fehlt, kann diese nicht der entsprechenden Klasse zugeordnet werden und somit auch nicht im VSB genutzt werden. Zur Zeit ist im VSB der Umgang mit Interferenzen auch schwierig gestaltet. Liegen

zum Beispiel in einer als *Zahl* erkannten Eigenschaft auch *Datums*-Werte vor, ist es nicht möglich Filter für beide Fälle zu definieren. Des Weiteren könnte eine teilweise Übersetzung der Bezeichner und Beschreibungen für die Nutzer verwirrend sein, da die Elemente in Ergebnislisten eventuell in unterschiedlichen Sprachen angezeigt werden.

Da der Visual SPARQL Builder als allgemeines Werkzeug programmiert ist, kann er theoretisch mit jeder semantischen Wissensbasis benutzt werden, die einen SPARQL-Endpunkt für den Datenzugriff anbietet. Da die Installation des VSB denkbar einfach ist, bietet es sich an, ihn an der eigenen Wissensbasis auszuprobieren.

Literaturverzeichnis

[Ob11] Object Management Group, Unified Modeling Language Infrastructure Version 2.4.1, http://www.omg.org/spec/UML/2.4.1/Infrastructure/PDF, Stand: 13.04.2015.

[Ru08] Russell, Alistair; Smart, Paul R.; Braines, Dave; Shadbolt, Nigel R.: NITELIGHT: A Graphical Tool for Semantic Query Construction. In: Semantic Web User Interaction Workshop (SWUI 2008). April 2008. Event Dates: 5th April 2008.

[SWL14] Stadler, Claus; Westphal, Patrick; Lehmann, Jens: Jassa - A JavaScript Suite for SPARQL-based Faceted Search. In: Proceedings of the ISWC Developers Workshop 2014. S. 31–36, 2014.

[Wo04] World Wide Web Consortium, OWL Web Ontology Language, http://www.w3.org/TR/2004/REC-owl-semantics-20040210/, Stand: 13.04.2015.

[Wo12] World Wide Web Consortium, XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes, Abschnitt Primitive Datatypes, http://www.w3.org/TR/2012/REC-xmlschema11-2-20120405/#built-in-primitive-datatypes, Stand: 13.04.2015.

[Wo13a] World Wide Web Consortium, SPARQL 1.1 Query Language, http://www.w3.org/TR/2013/REC-sparql11-query-20130321, Stand: 13.04.2015.

[Wo13b] World Wide Web Consortium, SPARQL 1.1 Query Language, Abschnitt Constraints in Optional Pattern Matching, http://www.w3.org/TR/2013/REC-sparql11-query-20130321/#OptionalAndConstraints, Stand: 13.04.2015.

[Wo13c] World Wide Web Consortium, SPARQL 1.1 Query Language, Abschnitt Expressions and Testing Values, http://www.w3.org/TR/2013/REC-sparql11-query-20130321/#expressions, Stand: 13.04.2015.

[Wo14a] World Wide Web Consortium, RDF 1.1 Semantics, http://www.w3.org/TR/2014/REC-rdf11-mt-20140225/, Stand: 13.04.2015.

[Wo14b] World Wide Web Consortium, RDF Schema 1.1, http://www.w3.org/TR/2014/REC-rdf-schema-20140225/, Stand: 13.04.2015.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Erzeugung minimaler Spannbäume auf ungerichteten, kantengewichteten Graphen mit den Algorithmen von Kruskal, Prim und Boruvka

Paul Jähne[1]

Abstract: Das Paper analysiert die Laufzeit der Algorithmen von Kruskal, Prim und Borůvka für das Erzeugen minimaler Spannbäume auf ungerichteten, kantengewichteten Graphen und vergleicht die Ergebnisse mit der theoretischen Laufzeit. Dazu werden die Algorithmen kurz vorgestellt und deren Implementierung erläutert. Die Algorithmen sind in der Programmiersprache C umgesetzt und teilweise mit dem MPI-Standard parallelisiert. Darauf folgt eine Beschreibung der für die Tests verwendeten Daten. Für Tests mit Graphen verschiedener Dichte werden generische Graphen fester Größe verwendet. Um Aussagen über Graphen verschiedener Größe zu treffen, kommen die Graphen des neunten DIMACS Implementierungswettbewerbs zum Einsatz. Anhand dieser Graphen wird gezeigt, dass der Algorithmus von Borůvka in der sequentiellen Variante sowohl für verschiedene Dichten als auch unterschiedliche Größen die schnellste Laufzeit und den geringsten Speicherverbrauch erzielt. Mit der gewählten Parallelisierungsstrategie erreicht der Algorithmus von Kruskal für dünne Graphen die beste Laufzeit.

Keywords: minimaler Spannbaum, Laufzeitanalyse, Algorithmus von Kruskal, Algorithmus von Prim, Algortihmus von Borůvka

1 Einleitung

Im Rahmen dieser Arbeit werden die Algorithmen von Kruskal, Prim und Borůvka für die Berechnung von minimalen Spannbäumen verglichen. Ein Spannbaum ist ein Baum aus Kanten eines Graphen, der alle Knoten enthält und kreisfrei ist. Minimal wird ein Spannbaum genannt, wenn die Summe aller Kantengewichte die kleinstmögliche unter allen Spannbäumen des Graphen darstellt.

Minimale Spannbäume werden unter anderem für Näherungslösungen des Rundreiseproblems oder für die Erstellung kostengünstiger Versorgungsnetzwerke eingesetzt. Sie sind auch zur Generierung von Labyrinthen verwendbar.

Es existieren bereits Veröffentlichungen, die die Algorithmen vergleichen. Jedoch sind diese relativ alt und stellen den Quellcode nicht zur Verfügung [MS91][DG98][BH01] oder benutzen einfache Implementierungen der Algorithmen [Po08]. Weiterhin fehlt bei diesen ein Bezug zu realen Daten, da mit zufällig generierten Graphen gearbeitet wird. Diese Lücke soll geschlossen werden. Zusätzlich wird mit den verwendeten Graphen ein größerer Problembereich abgedeckt.

[1] Hochschule für Technik, Wirtschaft und Kultur, Fakultät Informatik, Mathematik und Naturwissenschaftten, Schulstraße 12, 04687 Trebsen, paul.jaehne@gmx.de

2 Algorithmen

Alle verwendeten Verfahren gehören zur Klasse der Greedy-Algorithmen. Diese konstruieren die optimale Gesamtlösung aus Einzelentscheidungen, die zum jeweiligen Zeitpunkt den größten Gewinn versprechen. Bei der Angabe der asymptotischen Laufzeit steht E für die Menge der Kanten und V für die Knotenmenge des Graphen.

Der Algorithmus von Kruskal [Kr56] sucht auf dem gesamten Graphen nach der aktuell besten Lösung. Es handelt sich somit um ein globales Verfahren, das wie folgt abläuft. Zu Beginn erfolgt eine Sortierung der Kanten des Graphen nach ihrem Gewicht. Danach durchläuft der Algorithmus die sortierte Kantenliste. Dabei wird zu jeder Kante geprüft, ob ein Hinzufügen zum aktuellen Spannbaum einen Kreis entstehen lassen und somit die Spannbaumeigenschaft verletzen würde. Ist das nicht der Fall, wird die Kante zum Spannbaum hinzugefügt. Dies wird wiederholt bis alle Kanten abgearbeitet wurden. Um die Kreisfreiheit effizient zu prüfen, können die einzelnen Komponenten des Spannbaums in einer Union-Find-Struktur verwaltet werden. Unter Verwendung dieser wird die Laufzeit des Algorithmus durch das Sortieren der Kanten dominiert. Es ergibt sich somit für ein optimales vergleichsbasiertes Sortierverfahren die Zeitkomplexität $O(|E|\log|E|)$. Dies ist asymptotisch äquivalent zu $O(|E|\log|V|)$.

Im Gegensatz zum Algorithmus von Kruskal sucht der Algorithmus von Prim [Pr57] nur von den bisher gewählten Knoten aus eine neue Kante und arbeitet somit lokal. Am Anfang wird ein beliebiger Knoten gewählt. Danach wird die folgenden Schritte abgearbeitet. Von den verbundenen Knoten werden alle Kanten betrachtet, die zu einem bisher unerreichten Knoten führen. Von jenen wird die Kante mit dem geringsten Gewicht zum Spannbaum hinzugefügt. Dies wird fortgesetzt bis alle Knoten im Spannbaum enthalten sind. Für eine effiziente Implementierung wird eine Prioritätswarteschlange benötigt. Diese enthält zu jedem bisher unerreichten Knoten die Kante mit dem geringsten Gewicht, die zu einem erreichten Knoten führt. Weiterhin wird eine Adjazenzliste genutzt, in der die Nachbarn zu jedem Knoten enthalten sind. Für die Umsetzung der Warteschlange werden Heaps verwendet. Unter Nutzung eines binären Heap ergibt sich eine asymptotische Laufzeit von $O(|E|\log|V|)$ und mit einem Fibonacci-Heap $O(|E|+|V|\log|V|)$.

Der Algorithmus von Borůvka [Bo26] bearbeitet den Graphen folgendermaßen in einer Mischung aus lokalen und globalen Schritten. Beginne mit jedem Knoten als eine eigene Komponente. Finde von jeder Komponente die Kante mit dem geringsten Gewicht, die zu einer anderen Komponente führt. Füge diese Kanten hinzu und vereinige alle so verbundenen Komponenten. Wiederhole dies bis nur noch eine Komponente übrig ist. Somit wird auf dem gesamten Graphen für jede Komponente lokal nach dem aktuell besten Schritt gesucht. Für die Verwaltung der Komponenten kann eine Union-Find-Datenstruktur genutzt werden. Mit dieser ergibt sich eine theoretische Laufzeit von $O(|E|\log|V|)$.

3 Implementierung

Für die Implementierung der vorgestellten Algorithmen wird die Programmiersprache C verwendet. Weiterhin erfolgt eine teilweise Parallelisierung unter Verwendung des Messa-

ge Passing Interface (MPI)-Standard. Dabei handelt es sich um ein nachrichtengekoppeltes Programmiermodell. Dies bedeutet, dass jeder Prozess seinen eigenen Speicherbereich besitzt und zwischen diesen Informationen über Nachrichten ausgetauscht werden. Der Quellcode des Programms ist auf GitHub unter der LGPL in Version 3 verfügbar.[2]

Der Graph ist als C-struct mit einem Array für die Kantenliste und Werten für die Anzahl der Kanten und Knoten repräsentiert. Das Einlesen eines Graphen erfolgt über eine Datei in der zuerst die Anzahl der Knoten und Kanten steht. Danach folgen alle Kanten, die als drei Zahlen repräsentiert werden. Diese stehen für die verbundenen Knoten und das Kantengewicht. Die Knoten sind von Null beginnend durchnummeriert.

Die Implementierung des Algorithmus von Kruskal verwendet als Sortierverfahren Mergesort. Das Verfahren besitzt eine Zeitkomplexität von $O(n \log n)$ und ist ein stabiles Sortierverfahren. Weiterhin kann dieser Teil parallel ausgeführt werden, sodass jeder Prozess einen Teil der Kanten empfängt und sortiert. Anschließend werden die sortierten Teile wieder versendet und schrittweise vereinigt. Der Merge-Algorithmus ist leicht abgewandelt, indem die zweite Hälfte für das Merge in umgekehrter Reihenfolge kopiert wird. Das folgende Abarbeiten der Kantenliste erfolgt sequentiell unter Verwendung einer arraybasierten Union-Find-Struktur. Diese ist mit Pfadverkürzung und Union-by-Rank umgesetzt. Mit dieser wird geprüft, ob die Knoten der Kanten zu verschiedenen Komponenten gehören und somit ein kreisfreies Einfügen möglich ist.

Der Algorithmus von Prim wird mit Heaps umgesetzt. Es steht eine Variante mit einem binären Heap und eine mit einem Fibonacci-Heap zur Verfügung. Der binäre Heap ist als Array umgesetzt und der Fibonacci-Heap mit doppelt verketteten Listen. Außerdem wird eine Adjazenzliste benötigt, um auf die benachbarten Knoten zu jeder Kante zuzugreifen. Beide Varianten laufen gleichermaßen ab. Zuerst wird der Heap mit allen Knoten initialisiert. Begonnen wird beim ersten Knoten. Dieser wird entfernt und der Heap mit den Kanten aus der Adjazenzliste zu dem Knoten aktualisiert. Danach wird in jedem Schritt das Minimum entfernt und der Heap mit den neuen Kanten aktualisiert. Eine Parallelisierung erfolgt nicht, da die Heapstrukturen nicht dafür geeignet sind.

Der Algorithmus von Borůvka besteht aus zwei Teilen, die wiederholt nacheinander ausgeführt werden bis nur noch eine Komponente übrig ist.Der erste Teil ist die Suche aller Kanten mit dem geringstem Gewicht zwischen den Komponenten. Dazu wird die Union-Find-Struktur und ein Array von der Größe der Knotenanzahl genutzt. In dieses wird die Kante mit dem geringsten Kantengewicht eingetragen, die zu dem jeweiligen Knoten führt, sofern beide Knoten in verschiedenen Komponenten enthalten sind. Für die Ausführung mit mehreren Prozessoren wird die Suche parallelisiert. Dafür erhält jeder Prozess einen Teil der Kanten und sucht in diesem. Die Ergebnisse werden in einer Sammeloperation bei einem Prozess zusammengeführt. Dieser verteilt danach das Ergebnis wieder. Im zweiten Teil werden die neu verbundenen Komponenten vereinigt. Dazu durchläuft jeder Prozess das Array mit den kürzesten Knoten und vereinigt die verbundenen Komponenten. Anschließend aktualisiert jeder Prozess seine Komponentenliste.

[2] https://github.com/SethosII/minimum-spanning-tree

4 Daten

Um verschiedene Eigenschaften der Algorithmen zu untersuchen, werden zwei verschiedene Datenquellen verwendet. Mit den realen Daten werden die Eigenschaften für Graphen verschiedener Größe untersucht. Die generischen Daten dienen zur Untersuchung des Verhaltens auf Graphen unterschiedlicher Dichte.

Als Quelle für reale Daten werden die Graphen des 9. Implementierungswettbewerbs des Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) verwendet.[3] Die Graphen beschreiben das Straßennetz der USA bzw. Teile davon. Eine Auflistung der einzelnen Graphen findet sich in Tabelle 1.

Name	Beschreibung	Knoten	Kanten	Dichte
USA	USA komplett	23.947.347	58.333.344	2,03E-7
CTR	Zentral-USA	14.081.816	34.292.496	3,45E-7
W	West-USA	6.262.104	15.248.146	7,77E-7
E	Ost-USA	3.598.623	8.778.114	1,35E-6
LKS	Große Seen	2.758.119	6.885.658	1,81E-6
CAL	Kalifornien und Nevada	1.890.815	4.657.742	2,60E-6
NE	Nordost-USA	1.524.453	3.897.636	3,35E-6
NW	Nordwest-USA	1.207.945	2.840.208	3,89E-6
FLA	Florida	1.070.376	2.712.798	4,73E-6
COL	Colorado	435.666	1.057.066	1,11E-5
BAY	Bucht von San Francisco	321.270	800.172	1,55E-5
NY	New York City	264.346	733.846	2,10E-5

Tab. 1: Auflistung der realen Graphen

Die Graphen besitzen sehr unterschiedliche Größen und decken einen Bereich von 70.000 bis 60.000.000 Kanten ab. Alle Graphen sind dünn, da ihre Dichte zwischen 10^{-5} und 10^{-8} liegt. Die Dichte errechnet sich mit Formel 1.

$$Dichte(G) = \frac{2|E|}{|V|(|V|-1)} \approx \frac{2|E|}{|V|^2} \tag{1}$$

Für Tests während der Entwicklung und zum Vergleich der Algorithmen auf verschieden dichten Graphen werden generischen Graphen verwendet. Diese besitzen ein zufälliges Kantengewicht von 0 bis 99 und sind in Tabelle 2 aufgelistet. Alle haben 10.000 Knoten und unterscheiden sich in der Anzahl der Kanten. Der Gittergraph hat eine zweidimensionale, rasterartige Struktur. Außer am Rand hat jeder Knoten vier Nachbarn. Bei dem Pfadgraph handelt es sich um eine Kette von Knoten, die bis auf den Start- und Endknoten jeweils nur zwei Nachbarn haben. Somit hat dieser die kleinstmögliche Dichte für einen zusammenhängenden Graph.

Der Spannbaum des Gittergraphen kann als Labyrinth interpretiert werden. Das Programm verfügt über die Möglichkeit beliebige rechteckige Labyrinthe zu erzeugen.

[3] http://www.dis.uniroma1.it/challenge9/download.shtml

Name	Beschreibung	Knoten	Kanten	Dichte
FULL	Graph mit Dichte 1	10.000	49.995.000	1,00E0
HALF	Graph mit Dichte 0,5	10.000	24.997.500	5,00E-1
QUAR	Graph mit Dichte 0,25	10.000	12.498.750	2,50E-1
GRID	Gittergraph	10.000	19.800	3,96E-4
PATH	Pfadgraph	10.000	9.999	2,00E-4

Tab. 2: Auflistung der generischen Graphen

5 Laufzeitmessung

Die Laufzeitmessungen erfolgen auf einem Dell R720 Server mit zwei Intel Xeon E5-2680 v2 Prozessoren und 512 GB Arbeitsspeicher. Als Betriebssystem kommt das Community Enterprise Operating System (CentOS) in der Version 6.6 zum Einsatz. Dieses verwendet den Linux-Kernel der Version 2.6.32. Als Compiler dient der GNU C Compiler 4.8.1 und die MPI-Implementierung OpenMPI 1.8.2 für die Parallelisierung.

Es werden für jeden Algorithmus auf allen vorgestellten Daten jeweils 30 Messungen durchgeführt und über diese gemittelt. Für die parallelisierten Varianten werden die gleichen Messungen zusätzlich auf zwei, vier und acht Prozessorkernen ausgeführt. Die Zeitmessung erfolgt über die MPI-Funktion MPI_Wtime.

Abbildung 1 zeigt die sequentielle Laufzeit der Algorithmen auf Graphen unterschiedlicher Größe. Die beste Laufzeit für alle Größen liefert der Algorithmus von Borůvka. Die benötigte Zeit liegt in etwa bei der Hälfte im Vergleich zum Algorithmus von Kruskal. Die Algorithmen von Kruskal und Prim mit binärem Heap sind für kleinere Graphen gleich auf. Mit steigender Größe erreicht der Algorithmus von Kruskal bessere Laufzeiten. Der Algorithmus von Prim unter Verwendung eines Fibonacci-Heaps ist der langsamste. Es wird nur für größere Graphen eine bessere Laufzeit als mit einem binären Heap erzielt.

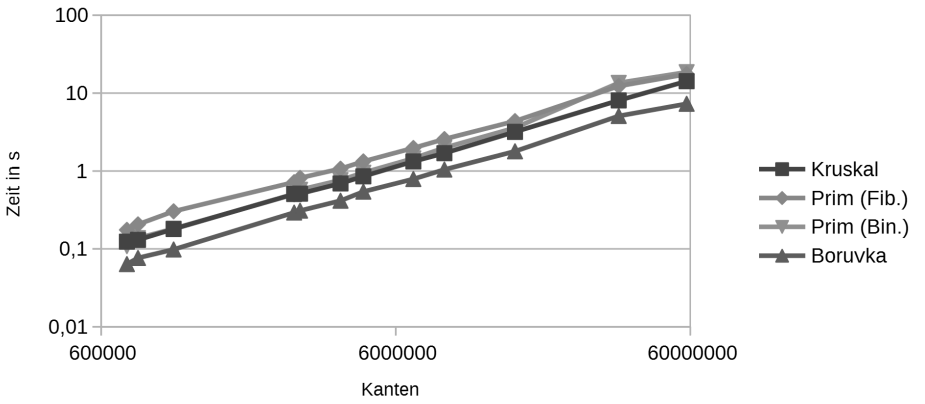


Abb. 1: Laufzeit für reale Daten (sequentiell)

Für den Vergleich der Algorithmen auf unterschiedlich dichten Graphen werden generische Graphen mit der gleichen Knotenanzahl verwendet. Die Ergebnisse sind in Abbildung 2 dargestellt. Bei diesen schneidet der Algorithmus von Kruskal mit zunehmender Dichte schlechter ab. Somit ist dieser Algorithmus für dichte Graphen wenig geeignet. Die Zunahme der Laufzeit für die Algorithmen von Kruskal und Prim ist nahezu linear. Jedoch ist der Anstieg des Algorithmus von Prim geringer. Die Variante mit einem Fibonacci-Heap erreicht keine bessere Laufzeit als die mit einem binären Heap. Der Algorithmus von Borůvka ist deutlich der schnellste. Dieser hat zusätzlich ab einer mittleren Dichte nur eine geringe Zunahme der Laufzeit je dichter der Graph wird. Damit ist der Algorithmus von Borůvka für dichte Graphen am besten geeignet.

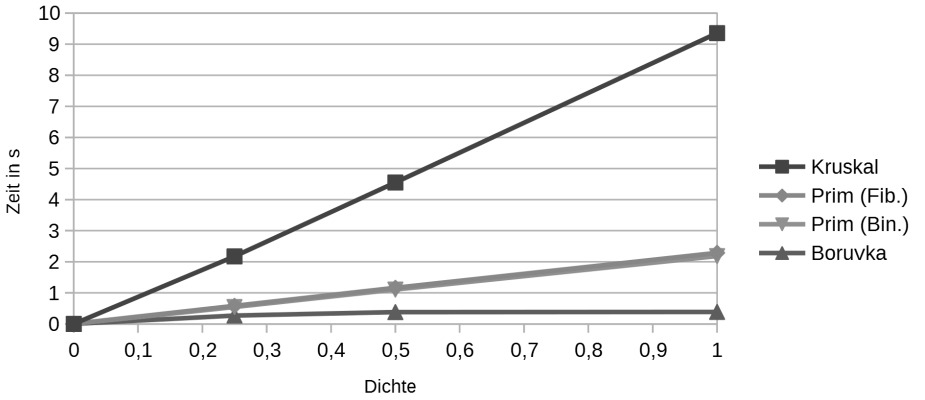


Abb. 2: Laufzeit in Abhängigkeit der Dichte (sequentiell)

Die Ergebnisse stehen im Gegensatz zu den Beobachtungen der eingangs angeführten Veröffentlichungen. So ergab die Veröffentlichung von Bazlamaçci und Hindi [BH01], dass der Algorithmus von Prim sowohl für dichte als auch dünne Graphen am besten geeignet ist und der Algorithmus von Borůvka mehr Zeit benötigt als der Algorithmus von Kruskal.

Im weiteren erfolgt eine Analyse der parallelen Varianten des Algorithmus von Kruskal und Borůvka für die realen und generischen Daten. Dazu wird der Speedup betrachtet. Dieser ergibt sich aus dem Verhältnis der Laufzeiten des sequentiellen und parallelen Algorithmus, wie in Formel 2 beschrieben, wobei P für die Anzahl der Prozessoren steht.

$$Speedup(P) = \frac{T(1)}{T(P)} \tag{2}$$

Die Abbildung 3 zeigt den erzielten Speedup auf Graphen verschiedener Größen unter Verwendung des parallelisierten Algorithmus von Kruskal. Dieser erreicht für alle verwendeten Größen der realen Daten einen akzeptablen Speedup, der unter Verwendung von mehr Prozessoren gesteigert werden kann. Für acht Prozessoren liegt dieser bei ungefähr 2,5. Mit vier und acht Prozessoren liegt die Laufzeit unter der sequentiellen Umsetzung des Algorithmus von Borůvka.

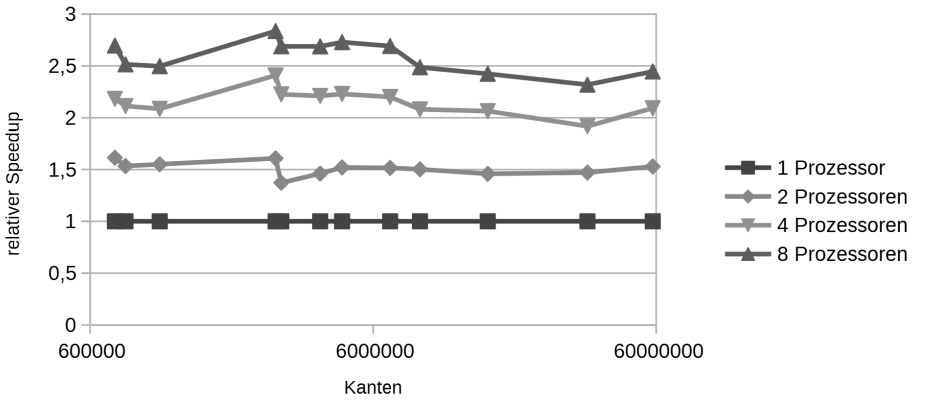


Abb. 3: Relativer Speedup für reale Daten (Kruskal)

Abbildung 4 zeigt den Speedup für den Algorithmus von Kruskal auf verschieden dichten Graphen. Für sehr dichte Graphen wird ein höherer Speedup erreicht als bei dünnen Graphen. Dies liegt daran, dass mit höherer Kantenanzahl der Sortieraufwand schneller steigt als der Kommunikationsaufwand. Unter Verwendung von acht Prozessoren kann ein Speedup nahe 4 erreicht werden. Damit liegt die Laufzeit für sehr dichte Graphen unter Verwendung von acht Prozessoren in etwa auf der Höhe des sequentiellen Algorithmus von Prim für diese Daten.

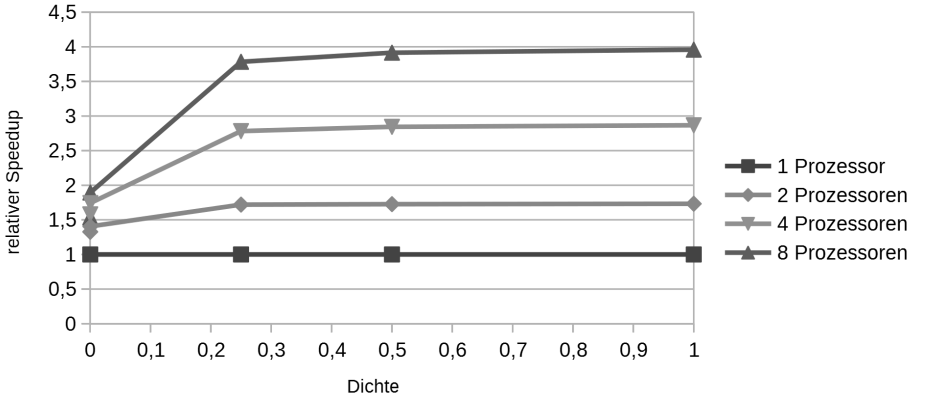


Abb. 4: Speedup in Abhängigkeit der Dichte (Kruskal)

In Abbildung 5 ist der Speedup für Graphen unterschiedlicher Größen unter Verwendung des Algorithmus von Borůvka dargestellt. Für reale Daten erreicht die verwendete Parallelisierung des Algorithmus von Borůvka keinen brauchbaren Speedup. Für zwei Prozessoren ist ein minimaler Zeitgewinn zu verzeichnen. Durch die Hinzunahme weiterer Prozessoren verlangsamt sich die Ausführung jedoch. Damit ist die gewählte Parallelisierungsstrategie für diese Daten nicht sinnvoll.

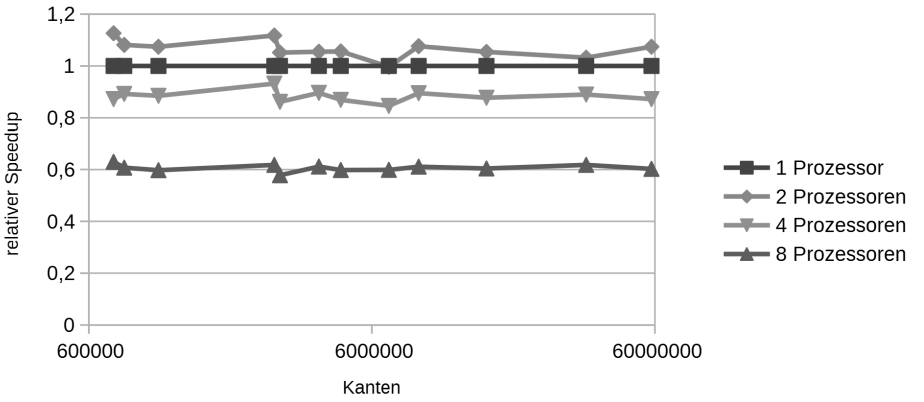


Abb. 5: Speedup für reale Daten (Borůvka)

Abbildung 6 zeigt den Speedup für den Algorithmus von Borůvka auf verschieden dichten Graphen. Für sehr dünne Graphen bestätigt sich das Bild aus den realen Daten, dass die Parallelisierung die Laufzeit verschlechtert. Für höhere Dichten ergibt sich jedoch ein akzeptabler Speedup, der mit steigender Dichte wieder sinkt. Somit kann die verwendete Parallelisierung für Graphen mittlerer Dichte als sinnvoll bezeichnet werden.

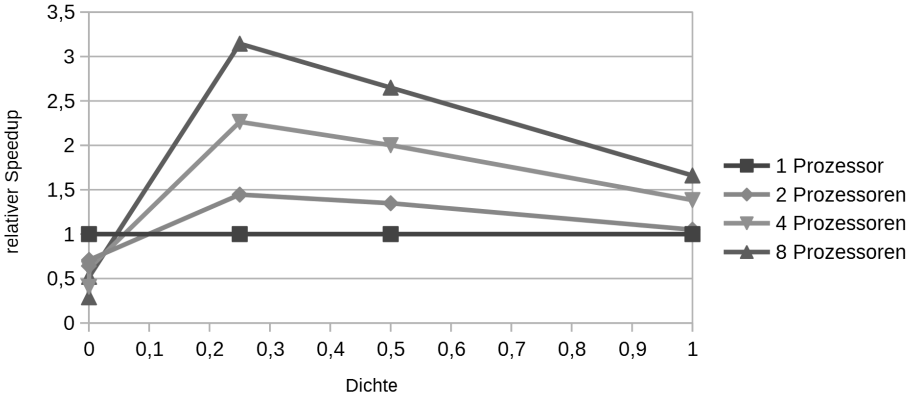


Abb. 6: Speedup in Abhängigkeit der Dichte (Borůvka)

6 Vergleich mit asymptotischer Laufzeit

Um zu überprüfen wie die reale Laufzeit zur asymptotischen in Beziehung steht, wird die Laufzeit der Algorithmen mit der jeweiligen asymptotischen Laufzeit aus Abschnitt 2 skaliert. Das Ergebnis ist in Abbildung 7 zu sehen.

Für alle Algorithmen ergibt sich eine waagerechte Linie. Dies bedeutet, dass die reale Laufzeit in dem Maße wächst, wie durch die asymptotische vorhergesagt. Die Implementierungen des Algorithmus von Prim enthalten Sprünge. Das zeigt an, dass eine leichte Abweichung von der asymptotischen Laufzeit vorliegt.

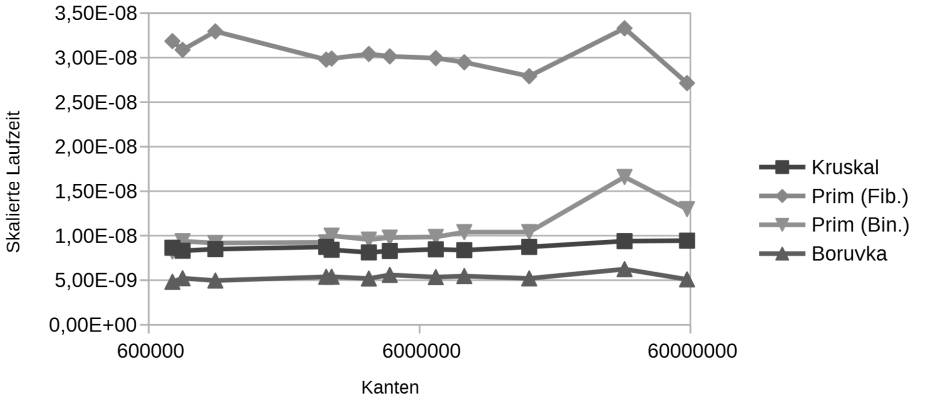


Abb. 7: Vergleich mit asymptotischer Laufzeit

7 Statistische Analyse

Um statistisch belegbare Aussagen für den Vergleich der Algorithmen treffen zu können, wird der gepaarte, zweiseitige t-Test verwendet. Die Resultate mit diesem Test stehen jeweils in Klammern hinter den Aussagen. Irrtumswahrscheinlichkeiten von unter einem Prozent (10^{-2}) sind als höchst signifikant anzusehen.

Anhand des USA-Graphen lassen sich folgende Aussagen treffen. Für die sequentiellen Varianten auf den realen Daten ist der Algorithmus von Borůvka schneller als der Algorithmus von Kruskal (10^{-32}). Dieser wiederum ist schneller als der Algorithmus von Prim in beiden Varianten (Prim Fib.: 10^{-26}, Prim Bin.: 10^{-23}). Der Algorithmus von Kruskal auf realen Daten ist in der parallelen Variante ab mehr als vier Prozessoren schneller als der Algorithmus von Borůvka (10^{-23}). Die Umsetzung des Algorithmus von Prim mit einem Fibonacci-Heap ist für große Graphen schneller als die mit einem binären Heap (10^{-17}). Die parallele Umsetzung des Algorithmus von Kruskal mit vier Prozessoren ist schneller als der sequentielle Algorithmus (10^{-23}).

Am Beispiel des FLA-Graphen kann gezeigt werden, dass die Laufzeiten der Algorithmen von Kruskal und Prim mit binärem Heap für kleine Graphen nicht verschieden sind (0,87). Außerdem kann gezeigt werden, dass der Algorithmus von Prim mit dem Fibonacci-Heap für kleine Graphen langsamer ist als der mit einem binären Heap (10^{-9}). Aus dem FULL-Graphen kann geschlossen werden, dass der Algorithmus von Borůvka für dichte Graphen schneller ist als der Algorithmus von Prim in beiden Varianten (Prim Fib.: 10^{-53}, Prim Bin.: 10^{-54}). Diese sind für dichte Graphen wiederum schneller als der Algorithmus von Kruskal (10^{-48}).

8 Speicherverbrauch

Für die Messung des Speicherverbrauchs wird der USA-Graph verwendet. Dies ist der größte verfügbare reale Graph. Die Messung erfolgt mit einem periodischen Aufruf des Kommandozeilentool ps aller zehn Millisekunden. Die Messwerte beschreiben den maximalen Speicherverbrauch über die gesamte Laufzeit der sequentiellen Varianten. Die Ergebnisse sind in Abbildung 8 zu sehen.

Die Repräsentation des Graphen benötigt im Arbeitsspeicher rund 0,7 GB, was als Referenzwert für die relative Angabe verwendet wird. Es ist zu erkennen, dass der Algorithmus von Borůvka den geringsten Speicherverbrauch hat (1,4 GB). Den meisten Speicher benötigt der Algorithmus von Prim unter Verwendung eines Fibonacci-Heaps (4,6 GB).

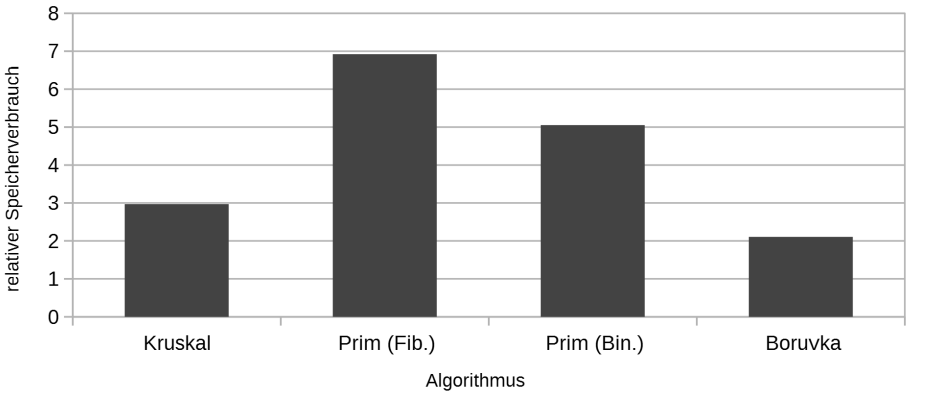


Abb. 8: Vergleich des relativen Speicherverbrauchs mit dem USA-Graphen

9 Zusammenfassung

Das Paper beschreibt und vergleicht die klassischen Algorithmen zur Berechnung eines minimalen Spannbaums. Es wird gezeigt, dass in der sequentiellen Variante der Algorithmus von Borůvka die beste Wahl ist. Dieser hat den geringsten Speicherverbrauch und die geringste Laufzeit für Graphen beliebiger Dichte und Größe. Zusätzlich ist er vergleichsweise einfach zu implementieren.

Die genutzte Parallelisierung des Algorithmus von Borůvka ist nicht geeignet für die Verwendung auf realen Daten und erreichte nur auf Graphen mit mittlerer Dichte einen akzeptablen Speedup. Mit einem speichergekoppelten Programmiermodell könnte die Parallelisierung sinnvoller umgesetzt werden, da der Austausch des Arrays nicht nötig wäre und stattdessen mit atomaren Operationen gearbeitet werden könnte.

Der Algorithmus von Kruskal ist für die realen Daten die zweitbeste Wahl. Durch die eingesetzte Parallelisierung kann mit genügend Prozessoren die Laufzeit des sequentiellen Algorithmus von Borůvka unterboten werden. Jedoch ist der Algorithmus von Kruskal für dichte Graphen ungeeignet.

Der Algorithmus von Prim wurde mit zwei verschiedenen Heaps untersucht. Dabei zeigt sich, dass der vergleichsweise einfache binäre Heap einem Fibonacci-Heap überlegen ist. Die einzige Ausnahme sind extrem große Graphen, bei denen der Fibonacci-Heap schneller ist. Zudem benötigt der Fibonacci-Heap den meisten Speicher und ist schwierig zu implementieren. Für kleine Graphen ist die Umsetzung mit dem binären Heap ähnlich schnell wie der Algorithmus von Kruskal. Bei sehr dichten Graphen stellt der Algorithmus von Prim die zweitbeste Wahl dar.

Insgesamt empfiehlt es sich daher für eine allgemeine Implementierung den Algorithmus von Borůvka zu verwenden.

Literaturverzeichnis

[BH01] Bazlamaçci, C. F.; Hindi, K. S.: Minimum-weight spanning tree algorithms - A survey and empirical study. Computers & Operations Research, 28:767–785, 2001.

[Bo26] Borůvka, O.: O jistém problému minimálním. Práca Moravské Přírodovědecké Společnosti, 3:37–58, 1926.

[DG98] Dehne, F.; Götz, S.: Practical Parallel Algorithms for Minimum Spanning Trees. IEEE Press, S. 366, 1998.

[Kr56] Kruskal, J. B.: On the shortest spanning subtree of a graph and the traveling salesman problem. Proceedings of the American Mathematical Society, 7:48–50, 1956.

[MS91] Moret, B. M. E.; Shapiro, H. D.: An Empirical Analysis of Algorithms for Constructing a Minimum Spanning Tree. In: DIMACS Series in Discrete Mathematics and Theoretical Computer Science. Springer, S. 400–411, 1991.

[Po08] Podsechin, Igor: , Comparing minimum spanning tree algorithms. http://www.aka.fi/Tiedostot/Tiedostot/Viksu/Viksu%202008/IgorPodsechin.pdf, 2008.

[Pr57] Prim, R. C.: Shortest connection networks and some generalizations. Bell System Technical Journal, 36:1389–1401, 1957.

Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

GIET: Generic Information Extraction using Triple Store Databases

Ronald Smith Djomkam Yotedje[1]

Abstract: Information extraction (IE) is a natural language processing (NLP) task which aims at retrieving specific type of information from natural language text. Before the actual extraction of information, a natural language text usually undergoes some processing, such as part-of-speech (POS) tagging. Nowadays, information extraction finds practical use in smart phones, which for example are able to identify a phone number in an SMS text and provide a link for either calling that number or storing it. There exist various approaches in extracting information from text, e.g., pattern-based information extraction which uses predefined rules to extract entities and ontology-based information extraction. These approaches often concentrate on a particular domain whose information is to be extracted. This paper provides a light-weight, unified, and scalable methodology which uses a generic approach for extracting information. This approach does not depend on any domain or rule.

1 Introduction

Information extraction (IE) is a technology based on analyzing natural language text with the aim of extracting specific type of information [Cu06, HDG00]. Depending on the system, the information usually extracted may be directly presented to the user requiring it, or saved in a database, or even used as a base for indexing purposes in information retrieval (IR) applications such as web-based search engines, e.g., Google and Duckduckgo. People sometimes tend to confuse between information extraction and retrieval. Compared to an information extraction system, an information retrieval system is a technology which acquires information resources that are related to the user's requirements. These information resources are often obtained from a large bank of information resources.

An example of an IE task is named entity recognition (NER). Named entity recognition searches and retrieves entities such as names, places, and dates. This is used in applications such as Google Mail, where dates are automatically recognized in texts and represented in such a way that one can easily create a calendar event from them.

There exists different approaches to IE, such as ontology-based information extraction, where unstructured or semi-structured natural language texts are processed through a mechanism guided by ontologies to extract certain types of information. The output is then presented using ontologies [WD10]. We present a light-weight, unified, and scalable methodology for IE, which makes use of triple store databases to store extracted information as triples. This provides a back-end architecture, which has the main function of combining various IE approaches, such as Named Entity Recognition and Co-reference

[1] Universität Bremen — FB3, AG Rechnerarchitektur, Bibliothekstrasse 1 (MZH), 28359 Bremen, smith@cs.uni-bremen.de

Resolution. We further use SPARQL as query language for extracting information through the back-ends.

The remainder of the paper is divided into five sections. Section 2 gives an overview of related work. Section 3 describes the main idea of this work followed by an algorithm. In Sections 4 and 5 we present respectively the system developed in this work and the experimental evaluation of this system, and finally, Section 6 gives an overall conclusion of the work.

2 Related work

Most of the fields related to our work has to do with natural language, linguistic annotation, and semantic web. Sonal Gupta [GM14] developed a system with two major functions. At first, it provides an information extraction system capable of extracting entities from a given text using patterns with bootstrapping. An advantage of this methodology is that the patterns used for entities extraction are not fixed but generated from the text. Further, the system provides a way of visualizing the extracted entities and patterns. The methodology described in [CJ11] is an amelioration of standard template-based information extraction, which uses predefined templates to fill semantic roles. This methodology removes the requirement of having a fix template, but instead introduces the learning of domains' template from the input text.

[ARR13] proposes a methodology for extracting information based on ontology, i.e., an ontology is used along to guide the extraction process. This methodology has a common point with the methodology proposed in this paper in that, it is generic, since it can be applied to any domain. [Ar13] proposes a methodology for automatic information extraction from natural language texts, based on the integration of linguistic rules, multiple ontologies and inference resources, integrated with an abstraction layer for linguistic annotation. [Ru12] proposes a way of representing linguistic annotation using the RDF framework.

NIF[3] and LemonWordNet[4] are both existing RDF wrappers for Standford Core NLP and WordNet respectively. Our methodology is not limited to a special nlp tool, as it allows someone to easily wrap an existing nlp tool for integration with other wrappers in the overall system. [KLA13] lets a user describe corpus-wide extraction tasks in a declarative language and permits him to run interactive rule refinement queries. The methodology used here has a similarity with ours, since the extraction actually takes place after issuing database queries. But the difference is that, we use a triple store database (TSDB) since it doesn't require a schema. This allows an easy customization of back-ends, which will be explained in detail in Section 4. Further, [KLA13] uses the declarative language AQL[5] for extraction. Our system uses the SPARQL RDF query language (SPARQL),[6] since the extracted information is stored in a TSDB.

[3] http://persistence.uni-leipzig.org/nlp2rdf
[4] http://wordnet-rdf.princeton.edu
[5] http://bigdataconsultants.blogspot.ch/2013/07/aql-annotation-query-language.html
[6] http://www.w3.org/TR/rdf-sparql-query

Mathias Soeken et al. [So14] proposed a triple store database information extraction method, which also uses the SPARQL query language for extraction. Our methodology proposes an improvement of this work with a back-end architecture, which enables different information extraction approaches to be combined.

3 Abstract concept

In this section we describe the main idea of our work and the algorithm used to develop the GIET system.

3.1 Main Idea

We propose a new methodology for extracting information from texts using triple store databases. This method is not only light-weight, but also unified and scalable. Light-weight is achieved by providing a back-end architecture with a simple interface that requires almost no dependencies. The method is also said to be unified since upon call of the simple interface from the user, all required back-ends are activated. The activated back-ends are responsible to retrieve the necessary information. Finally, the ability to extract just the necessary information as requested by the user makes the method scalable.

Example 1: Consider the following sentence: "Smith goes to Africa occasionally." The task is to retrieve named entities, and their types from the sentence. Further consider we have three different back-ends available. One back-end extracts words, another extracts named entities and the last one extracts part-of-speech tags from the sentence. The SPARQL query looks like the following:

```
SELECT ?word ?namedEntity
WHERE {
   ?w word ?word.
   ?w is ?namedEntity.
}
```

Each predicate of the query statement represents the type of information needed. There are two query statements. The first statement consists of a subject which has a word and the second statement says that this same subject should be a named entity. The word and the named entity will be saved respectively in the variables *word* and *namedEntity*. Since just named entities and their types are needed, the back-end extracting part-of-speech tags will not be considered anymore from the interface. The two other back-ends now retrieve corresponding information from the sentence and save them as triple in the triple store database. The following shows the triples stored in the database:

```
<goes-2> word "goes"
```

```
<Africa-4> word "Africa"
<Africa-4> is "LOCATION"
<Smith-1> word "Smith"
<Smith-1> is "PERSON"
<to-3> word "to"
<occasionally-5> word "occasionally"
```

Notice that each word in the sentence has a corresponding word-triple. This is very important, as a sentence may contain the same word more than one time. Applying the above query to the above triple store database yields the following query result:

?namedEntity	?type
Africa	LOCATION
Smith	PERSON

Tab. 1: Query result for example 1

Figure 1 gives an abstract overview of how this method is used to extract information. As observed in the picture, inputs are a sentence as information source and a query. The simple interface executes an algorithm as described in the next section. In this algorithm, we determine from the query the type of information to be retrieved, and which back-ends to involve for retrieving this information. The back-ends to be involved for retrieval depends on the type of information determined from the query. A back-end aims at retrieving information from the input sentence. There are no restrictions to how or which information is extracted from the sentence. This means back-ends can retrieve information as is, or enrich this information by using external dependencies. The advantage is that the simple interface does not know about the existence of any external dependencies, which makes it less coupled to the back-ends.

Figure 1 also shows a triple store database, where information is stored as triple after being retrieved from the sentence. Afterwards, the query gets executed on the triple store database to yield the requested information from the user.

3.2 Algorithm

This section describes the algorithm executed in the simple interface of the back-end architecture. As mentioned in the previous section, inputs to the algorithm are a sentence as information source and a query. Algorithm 1 shows the pseudo code of the algorithm.

The first step of the algorithm consists of determining the type of information from the query. As a next step, iteration takes place over the list of available back-ends where only those back-ends are saved which can retrieve the information as determined from the query. Further, each required back-end retrieves information from the sentence and stores it as triple in a triple store database. A triple represents the kind of relation which exists between two entities. Since the database used to store information is a triple store database, SPARQL was chosen as query language for extraction.

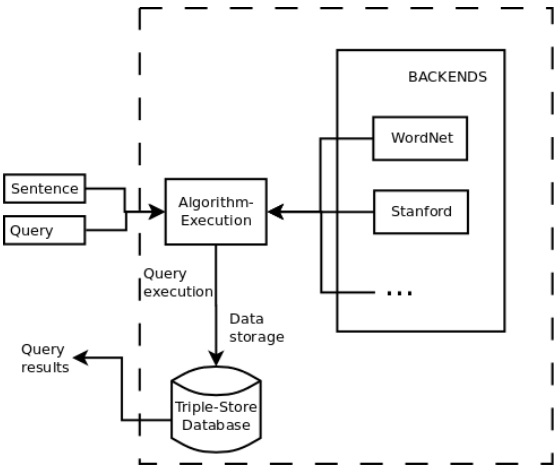


Fig. 1: Abstract overview of information extraction using our methodology

Algorithm 1

Input: $sentence \in String$
Input: $query \in String$
Output: $result \in QueryResult$
1: $informationType \leftarrow determineInfosType(query)$
2: $requiredBackends \leftarrow \{\}$
3: **for** $backend \in availableBackends()$ **do**
4: **if** $canRetrieveInfosOfType(backend, informationType)$ **then**
5: $requiredBackends \leftarrow requiredBackends \cup backend$
6: **end if**
7: **end for**
8: $listOfTriples \leftarrow \{\}$
9: **for** $backend \in requiredBackends$ **do**
10: $triples \leftarrow extractInformation(backend, sentence)$
11: $listOfTriples \leftarrow listOfTriples \cup triples$
12: **end for**
13: $store \leftarrow storeInDatabase(listOfTriples)$
14: $result \leftarrow execute(query, store)$

4 GIET system

As seen in Section 3, our main idea is to provide a more generic way of extracting information from a text. Below we present some freely available tools used.

- **Application service:** The application service[7] is a scalable tool realized in the Java programming language, which helps to develop client-server applications using the transmission control protocol (TCP) with very less effort. In using this tool, one has

[7] https://github.com/rosmith/application-service

to concentrate only on the logic of the application to be developed. Everything else like the network communication is handled by the application service, provided it is appropriately configured.

- **Stanford CoreNLP:** CoreNLP from the Stanford university is a Java annotation pipeline framework, which provides most of the common NLP steps, from tokenization through to co-reference resolution [Ma14].
- **Apache Jena:** Apache Jena[8] is a free and open source Java framework for building semantic web and linked data. The main goal of this tool is to provide facilities for both the application and the system programmers. The application programmer can access and manipulate graph data through higher level interfaces whereas a system programmer wishing to manipulate data as triples has a simple minimalist view of RDF graphs [Ca04].
- **Clear NLP:** This is another Java framework for NLP, which was developed by Jinho [CP11]. This NLP tool has almost same components as the Stanford CoreNLP, like part-of-speech (POS) tagging and dependency parsing, just that both use different approaches. Clear NLP provides an additional component known as semantic role labeling (SRL) which is not yet provided by the Stanford CoreNLP.
- **Extended Java WordNet Library (ExtJWNL):** WordNet is an online lexical database designed for use under program control. This database links English nouns, verbs, adjectives, and adverbs to sets of synonyms that are in turn linked through semantic relations that determine word definitions. WordNet contains more than 166,000 word form and sense pairs, where a word form is represented by a string of ASCII characters, and a sense is represented by the set of (one or more) synonyms that have that sense [Mi95]. ExtJWNL [9] is a free and open source Java framework for creating, reading and updating dictionaries in WordNet format.
- **Behind The Name (BTN) & Jsoup:** BTN[10] is a website which on request, gives the etymology of a name, which can be either a person's name or surname or even a place's name. Unfortunately, there is no existing library at the moment which enables one to use the functionalities of BTN programmatically. For that reason, one can make use of the library Jsoup,[11] which is a Java library for working with real-world HTML, to navigate in the HTML tree of a search result page and get some relevant information like the gender or description of the searched name.

The GIET system is implemented in Scala[12]. The source code is available in GitHub[13] under the MIT License, and a documentation can be found in the webpage[14], with an

[8] https://jena.apache.org/documentation/inference/
[9] http://extjwnl.sourceforge.net/
[10] http://www.behindthename.com
[11] http://jsoup.org
[12] http://scala-lang.org
[13] https://github.com/rosmith/giet
[14] http://giet-nlp.de

interactive user interface to play around with. Fig 2 shows the workflow of information extraction of the GIET system.

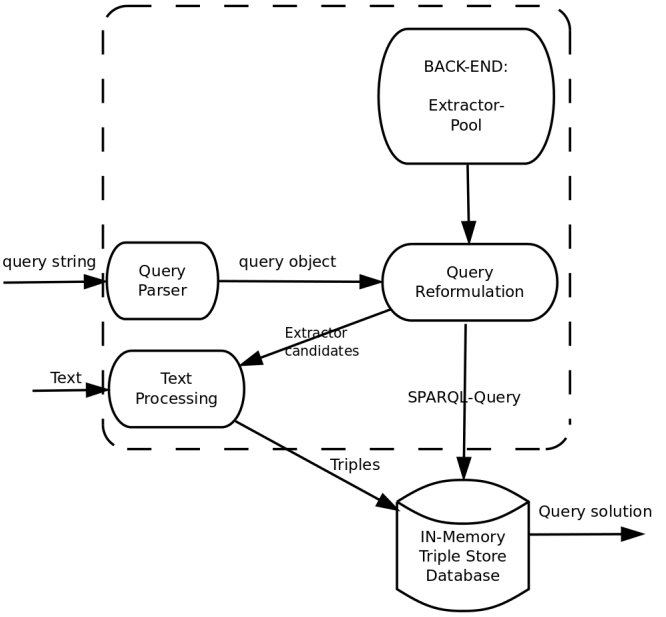


Fig. 2: Information extraction workflow of the GIET system

The GIET system provides a scalable back-end architecture. This allows combining different information extraction approaches to form a unified IE system. The back-end architecture consists of a pool of built-in extractors. An available extractor interface allows an application programmer to provide a custom information extractor, which retrieves information according to the needs of the application programmer. It should be emphasized that annotations can be represented by both data and object properties. According to the task done by an extractor, the programmer decides which representation best suites. Following are some important built-in extractors in the GIET system with their corresponding namespaces in brackets:

- **Word extractor (word)** wraps the Stanford tokenizer[15] and aims at saving token annotations of the sentence.
- **Tagger extractor (tag)** wraps the Stanford tagger [To03] and aims at saving part-of-speech tag annotations for each token in the sentence.
- **Dependencies extractor (dependency)** wraps the Stanford parser [KM03] and aims at saving typed dependencies annotations such as nominal subject (nsubj), adjective (jj), etc.
- **Coreference extractor (coref)** wraps the Stanford co-reference resolution system [Le11] and aims at saving existing co-reference annotations in the text.

[15] http://nlp.stanford.edu/software/tokenizer.shtml

- **Named entity extractor (is)** wraps the Stanford named entity recognition system [FGM05] and aims at saving named entity annotations.
- **SRL extractor (srl)** wraps the SRL component of the ClearNLP system [FGM05] and aims at saving semantic role annotations.
- **BTN extractor (btn)** uses both BTN and Jsoup to save people's gender and name definition. This shows that the GIET system is very flexible and allows someone to add new information, which is not present in the sentence.
- **WordNet extractor (wordnet)** uses the extended JWNL to save lexical fields information for the syntactic categories nouns, verbs, adjective, and adverbs. An example of a lexical field[16] for a verb is *motion* which contains verbs walking, flying and swimming.

All of the above extractors are available in the extractor pool at the beginning of the extraction. Each extractor has a unique namespace, with which it can be addressed. Extraction of information using the GIET system occurs as follow:

- The query string has a very simple format which abstracts the SPARQL query language. This query string is first parsed to obtain a query object. A query of all tokens in a sentence will look like the following:

```
SELECT ?token
WHERE {
   ?x word ?token.
}
```

  In the above query string, one can observe that the triple statement has a predicate without namespace. The predicate itself is a namespace and hence corresponds to the extractor with the namespace *word* which is the *WordExtractor*, and will be the only extractor retained for extraction, the rest will be kept uninstantiated in the pool.

- The query object helps to infer which extractors (built-in and custom) from the pool are needed for extraction. After resolving the needed extractors, the query object is reformulated to the following full SPARQL query.

```
PREFIX giet: <http://www.giet-nlp.de/annotator#>

SELECT ?token
WHERE {
  ?x giet:word ?token.
}
```

- The resolved extractor candidates are then used to retrieve information from the text. The resulting triples are then saved in the in-memory triple store database which is provided by the apache jena API as described above.
- The generated SPARQL query is then applied on the triple store database to extract the required information.

[16] https://wordnet.princeton.edu/man/lexnames.5WN.html

5 Use case description

For the use cases, we used the mixed case Test Data 4 provided by the Message Understanding Conference (MUC-4)[17]. The MUC was initiated by the former Naval Ocean Systems Center (NOSC) and financed by the Defense Advanced Research Projects Agency (DARPA) with the aim of encouraging the development of new methods for extracting information [GS96]. The test data consists of one hundred texts, all talking about terrorist activities in Latin Amerika.

Example 2: To illustrate some important use cases, we make use of the following two sentences from the second text of the test data.

> *The prosecutor of the military court trying the people responsible for the killing of 124 inmates accused of terrorist activities held in the Lurigancho Prison on 19 June 1986 has asked for 25 years imprisonment for the person responsible for the killing.*
>
> *Prosecutor Juan Carbone Herrera requested the 25 years imprisonment for General Rolando Cabezas Alarcon of the Republican Guard "for ordering the shooting of 124 of the San Pedro prison inmates (Lurigancho)."*

The above sentences contain two named entities of type person, two of type location, one of type date, and two others of type number and duration, which are actually referring to the same entity, i.e. 124 and 25 years, respectively. With the following query, we aim at extracting any entity being dependent of a verb. We are also interested in the verb governing this entity, the part-of-speech of both the entity and its' governor (verb), the type of entity it is, and the position of both in the sentences. According to the text, we expect to have one result.

```
select ?dependent ?governor ?namedEntity ?pos ?gpos ?tag ?gtag
where {
  // nsubj relates a verb and its subject.
  ?x nsubj ?y. // nsubj acts as an object property
  ?x word ?governor.
  ?y word ?dependent.
  ?y is ?namedEntity.
  ?x position ?gpos.
  ?y position ?pos.
  ?x tag ?gtag.
  ?y tag ?tag.
}
```

It should be noted that, the user is the one providing the sentence and the query. The query mostly depends on what the user needs to extract from the sentence. After giving

[17] http://www-nlpir.nist.gov/related_projects/muc/index.html

the above sentence and user-defined query to the system, the system creates a query object from the above query string, and infers from it, which of the extractors in the pool should be instantiated for retrieving the required information.

In our case, we have eight query statements with five different namespaces from different extractors. These extractors will retrieve all corresponding information from the sentences and save them in the triple store database. After applying the query on the database, we receive exactly one result as expected:

?dependent	?governor	?namedEntity	?pos	?gpos	?tag	?gtag
Herrera	requested	PERSON	4	5	NNP	VBD

Tab. 2: Query result for example 2

Further constraints could be applied, such as restricting the named entities to be of type location, which will of course have no result, or restricting the governor to belong to the lexical field *communication*, e.g.:

```
select ?dependent ?governor ?namedEntity ?pos ?gpos ?tag ?gtag
where {
  ?x nsubj ?y.
  ?x word ?governor.
  ?y word ?dependent.
  ?y is ?namedEntity.
  ?x position ?gpos.
  ?y position ?pos.
  ?x tag ?gtag.
  ?y tag ?tag.
  ?w communication ?d.
  ?w word ?c.
  FILTER (?governor = ?c)
}
```

Due to the three last lines of the above query, the WordNet extractor will be instantiated to retrieve words belonging to the lexical field *communication* and the filter makes sure that there is at least one word belonging to the lexical field *communication* which is the same as the governor.

Our system relies much on external dependencies, such as Stanford CoreNLP and WordNet, which have been used to implement the built-in extractors. This makes the quality of the query results to be strongly dependent on these external dependencies. In the above query, we could further constrain the query, such that the dependent have the gender male. However, there will be no result with this additional constraint, since the dependent in the sentence is *Herrera*, which is not available in the database of Behind The Name, used by the BTN extractor.

6 Conclusion

In this paper, we provided a generic methodology for information extraction. This methodology provides a scalable back-end architecture, which enables to combine different approaches for extraction. The methodology is light-weight due to the unawareness of the eventual dependencies used in the back-end architecture and the fact that with a single query, different information can be extracted, also makes the methodology unified.

In future work, we plan to develop a declarative domain specific language (DSL) for querying. The motivation is that, not every programmer knows what is SPARQL and how to use it, so a DSL will give the chance to these programmers to make use of the advantages of our system, without having to know how a SPARQL query works. We also intend to output the results of our system according to state-of-the-art formats for representation of linguistic annotation like NIF.

Acknowledgement

Allow me to dedicate my acknowledgment of gratitude toward the following significant advisors and contributors:

First and foremost, I would like to thank Dr. Mathias Soeken for his support and encouragement. He kindly read my paper and offered invaluable detailed advices on grammar, organization, and the theme of the paper. Second, I would like to thank Prof. Dr. Rolf Drechsler and Prof. Dr. phil. John A. Bateman to read my paper and to provide valuable advices.

References

[Ar13] de Araujo, Denis A.; Rigo, Sandro J.; Muller, Carolina; Chishman, Rove: Exploring the inference role in automatic information extraction from texts. In: The Joint Workshop on NLP & LOD and SWAIE. pp. 33–40, September 2013.

[ARR13] Anantharangachar, Raghu; Ramani, Srinivasan; Rajagopalan, S: Ontology Guided Information Extraction from Unstructured Text. International Journal of Web & Semantic Technology (IJWesT), 4(1):19, January 2013.

[Ca04] Carroll, Jeremy J.; Dickinson, Ian; Dollin, Chris; Reynolds, Dave; Seaborne, Andy; Wilkinson, Kevin: Jena: Implementing the Semantic Web Recommendations. In: The Thirteenth International World Wide Web Conference. Association for Computing Machinery (ACM), New York, New York, pp. 74–83, May 2004.

[CJ11] Chambers, Nathanael; Jurafsky, Dan: Template-Based Information Extraction without the Templates. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL). pp. 976–986, June 2011.

[CP11] Choi, Jinho D.; Palmer, Marther: Transition-based Semantic Role Labeling Using Predicate Argument Clustering. In: Proceedings of the ACL 2011 Workshop on Relational Models of Semantics (RELMS 2011). 2011.

[Cu06] Cunningham, H: Information Extraction, Automatic. Encyclopedia of Language & Linguistics, 5:665–677, 2006.

[FGM05] Finkel, Jenny Rose; Grenager, Trond; Manning, Christopher D.: Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In: Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005). pp. 363–370, 2005.

[GM14] Gupta, Sonal; Manning, Christopher D.: SPIED: Stanford Pattern-based Information Extraction and Diagnostics. In: Proceedings of the ACL 2014 Workshop on Interactive Language Learning, Visualization, and Interfaces (ACL-ILLVI). 2014.

[GS96] Grishman, Ralph; Sundheim, Beth: Message Understanding Conference: A Brief History. MUC-6, pp. 466–471, 1996.

[HDG00] Humphreys, Kevin; Demetriou, George; Gaizauskas, Robert: Two Applications of Information Extraction to Biological Science Journal Articles: Enzyme Interactions and Protein Structures. In: Pacific Symposium on Biocomputing. pp. 490–501, 2000.

[KLA13] Killias, Torsten; Loeser, Alexander; Andritsos, Periklis: INDREX: In-Database Distributional Relation Extraction. In: ACM 6th International Workshop on Data Warehousing and OLAP (DOLAP). 2013.

[KM03] Klein, Dan; Manning, Christopher D.: Accurate Unlexicalized Parsing. In: Proceedings of the 41st Meeting of the Association for Computational Linguistics. pp. 423–430, 2003.

[Le11] Lee, Heeyoung; Peirsman, Yves; Chang, Angel; Chambers, Nathanael; Surdeanu, Mihai; Jurafsky, Dan: Stanford's Multi-Pass Sieve Coreference Resolution System at the CoNLL-2011 Shared Task. In: Proceedings of the CoNLL-2011 Shared Task. 2011.

[Ma14] Manning, Christopher D.; Surdeanu, Mihai; Bauer, John; Finkel, Jenny; Bethard, Steven J.; McClosky, David: The Stanford CoreNLP Natural Language Processing Toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Association for Computational Linguistics, Baltimore, Maryland, pp. 55–60, June 2014.

[Mi95] Miller, George A.: WordNet: A Lexical Database for English. COMMUNICATIONS OF THE ACM, 38(11):39–41, November 1995.

[Ru12] Rubiera, Emilio; Polo, Luis; Berrueta, Diego; El Ghali, Adil: TELIX: An RDF-Based Model for Linguistic Annotation. In (Simperl, Elena; Cimiano, Philipp; Polleres, Axel; Corcho, Oscar; Presutti, Valentina, eds): The Semantic Web: Research and Applications, volume 7295 of Lecture Notes in Computer Science, pp. 195–209. Springer Berlin Heidelberg, 2012.

[So14] Soeken, Mathias; Harris, Christopher B.; Abdessaied, Nabila; Harris, Ian G.; Drechsler, Rolf: Automating the Translation of Assertions Using Natural Language Processing Techniques. In: Forum on Specification and Design Languages - Proceedings. o.A., 2014.

[To03] Toutanova, Kristina; Manning, Christopher D.; Klein, Dan; Singer, Yoram: Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In: Proceedings of HLT-NAACL. pp. 252–259, 2003.

[WD10] Wimalasuriya, Daya C.; Dou, Dejing: Ontology-based information extraction: An introduction and a survey of current approaches. Journal of Information Science, 36(3):306–323, 2010.

GI-Edition Lecture Notes in Informatics

P-1 Gregor Engels, Andreas Oberweis, Albert Zündorf (Hrsg.): Modellierung 2001.

P-2 Mikhail Godlevsky, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications, ISTA'2001.

P-3 Ana M. Moreno, Reind P. van de Riet (Hrsg.): Applications of Natural Lan-guage to Information Systems, NLDB'2001.

P-4 H. Wörn, J. Mühling, C. Vahl, H.-P. Meinzer (Hrsg.): Rechner- und sensorgestützte Chirurgie; Workshop des SFB 414.

P-5 Andy Schürr (Hg.): OMER – Object-Oriented Modeling of Embedded Real-Time Systems.

P-6 Hans-Jürgen Appelrath, Rolf Beyer, Uwe Marquardt, Heinrich C. Mayr, Claudia Steinberger (Hrsg.): Unternehmen Hochschule, UH'2001.

P-7 Andy Evans, Robert France, Ana Moreira, Bernhard Rumpe (Hrsg.): Practical UML-Based Rigorous Development Methods – Countering or Integrating the extremists, pUML'2001.

P-8 Reinhard Keil-Slawik, Johannes Magenheim (Hrsg.): Informatikunterricht und Medienbildung, INFOS'2001.

P-9 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Innovative Anwendungen in Kommunikationsnetzen, 15. DFN Arbeitstagung.

P-10 Mirjam Minor, Steffen Staab (Hrsg.): 1st German Workshop on Experience Management: Sharing Experiences about the Sharing Experience.

P-11 Michael Weber, Frank Kargl (Hrsg.): Mobile Ad-Hoc Netzwerke, WMAN 2002.

P-12 Martin Glinz, Günther Müller-Luschnat (Hrsg.): Modellierung 2002.

P-13 Jan von Knop, Peter Schirmbacher and Viljan Mahni_ (Hrsg.): The Changing Universities – The Role of Technology.

P-14 Robert Tolksdorf, Rainer Eckstein (Hrsg.): XML-Technologien für das Semantic Web – XSW 2002.

P-15 Hans-Bernd Bludau, Andreas Koop (Hrsg.): Mobile Computing in Medicine.

P-16 J. Felix Hampe, Gerhard Schwabe (Hrsg.): Mobile and Collaborative Business 2002.

P-17 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Zukunft der Netze –Die Verletzbarkeit meistern, 16. DFN Arbeitstagung.

P-18 Elmar J. Sinz, Markus Plaha (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2002.

P-19 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund.

P-20 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund (Ergänzungsband).

P-21 Jörg Desel, Mathias Weske (Hrsg.): Promise 2002: Prozessorientierte Methoden und Werkzeuge für die Entwicklung von Informationssystemen.

P-22 Sigrid Schubert, Johannes Magenheim, Peter Hubwieser, Torsten Brinda (Hrsg.): Forschungsbeiträge zur "Didaktik der Informatik" – Theorie, Praxis, Evaluation.

P-23 Thorsten Spitta, Jens Borchers, Harry M. Sneed (Hrsg.): Software Management 2002 – Fortschritt durch Beständigkeit

P-24 Rainer Eckstein, Robert Tolksdorf (Hrsg.): XMIDX 2003 – XML-Technologien für Middleware – Middleware für XML-Anwendungen

P-25 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Commerce – Anwendungen und Perspektiven – 3. Workshop Mobile Commerce, Universität Augsburg, 04.02.2003

P-26 Gerhard Weikum, Harald Schöning, Erhard Rahm (Hrsg.): BTW 2003: Datenbanksysteme für Business, Technologie und Web

P-27 Michael Kroll, Hans-Gerd Lipinski, Kay Melzer (Hrsg.): Mobiles Computing in der Medizin

P-28 Ulrich Reimer, Andreas Abecker, Steffen Staab, Gerd Stumme (Hrsg.): WM 2003: Professionelles Wissensmanagement – Er-fahrungen und Visionen

P-29 Antje Düsterhöft, Bernhard Thalheim (Eds.): NLDB'2003: Natural Language Processing and Information Systems

P-30 Mikhail Godlevsky, Stephen Liddle, Heinrich C. Mayr (Eds.): Information Systems Technology and its Applications

P-31 Arslan Brömme, Christoph Busch (Eds.): BIOSIG 2003: Biometrics and Electronic Signatures

P-32 Peter Hubwieser (Hrsg.): Informatische Fachkonzepte im Unterricht – INFOS 2003

P-33 Andreas Geyer-Schulz, Alfred Taudes (Hrsg.): Informationswirtschaft: Ein Sektor mit Zukunft

P-34 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)

P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)

P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenberg (Hrsg.): Informatik 2003 – Mit Sicherheit Informatik

P-37 Arndt Bode, Jörg Desel, Sabine Rathmayer, Martin Wessner (Hrsg.): DeLFI 2003: e-Learning Fachtagung Informatik

P-38 E.J. Sinz, M. Plaha, P. Neckel (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2003

P-39 Jens Nedon, Sandra Frings, Oliver Göbel (Hrsg.): IT-Incident Management & IT-Forensics – IMF 2003

P-40 Michael Rebstock (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2004

P-41 Uwe Brinkschulte, Jürgen Becker, Dietmar Fey, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle, Thomas Runkler (Edts.): ARCS 2004 – Organic and Pervasive Computing

P-42 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Economy – Transaktionen und Prozesse, Anwendungen und Dienste

P-43 Birgitta König-Ries, Michael Klein, Philipp Obreiter (Hrsg.): Persistance, Scalability, Transactions – Database Mechanisms for Mobile Applications

P-44 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): Security, E-Learning. E-Services

P-45 Bernhard Rumpe, Wofgang Hesse (Hrsg.): Modellierung 2004

P-46 Ulrich Flegel, Michael Meier (Hrsg.): Detection of Intrusions of Malware & Vulnerability Assessment

P-47 Alexander Prosser, Robert Krimmer (Hrsg.): Electronic Voting in Europe – Technology, Law, Politics and Society

P-48 Anatoly Doroshenko, Terry Halpin, Stephen W. Liddle, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications

P-49 G. Schiefer, P. Wagner, M. Morgenstern, U. Rickert (Hrsg.): Integration und Datensicherheit – Anforderungen, Konflikte und Perspektiven

P-50 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 1) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm

P-51 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 2) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm

P-52 Gregor Engels, Silke Seehusen (Hrsg.): DELFI 2004 – Tagungsband der 2. e-Learning Fachtagung Informatik

P-53 Robert Giegerich, Jens Stoye (Hrsg.): German Conference on Bioinformatics – GCB 2004

P-54 Jens Borchers, Ralf Kneuper (Hrsg.): Softwaremanagement 2004 – Outsourcing und Integration

P-55 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): E-Science und Grid Adhoc-Netze Medienintegration

P-56 Fernand Feltz, Andreas Oberweis, Benoit Otjacques (Hrsg.): EMISA 2004 – Informationssysteme im E-Business und E-Government

P-57 Klaus Turowski (Hrsg.): Architekturen, Komponenten, Anwendungen

P-58 Sami Beydeda, Volker Gruhn, Johannes Mayer, Ralf Reussner, Franz Schweiggert (Hrsg.): Testing of Component-Based Systems and Software Quality

P-59 J. Felix Hampe, Franz Lehner, Key Pousttchi, Kai Ranneberg, Klaus Turowski (Hrsg.): Mobile Business – Processes, Platforms, Payments

P-60 Steffen Friedrich (Hrsg.): Unterrichtskonzepte für inforrmatische Bildung

P-61 Paul Müller, Reinhard Gotzhein, Jens B. Schmitt (Hrsg.): Kommunikation in verteilten Systemen

P-62 Federrath, Hannes (Hrsg.): „Sicherheit 2005“ – Sicherheit – Schutz und Zuverlässigkeit

P-63 Roland Kaschek, Heinrich C. Mayr, Stephen Liddle (Hrsg.): Information Systems – Technology and ist Applications

P-64 Peter Liggesmeyer, Klaus Pohl, Michael Goedicke (Hrsg.): Software Engineering 2005

P-65 Gottfried Vossen, Frank Leymann, Peter Lockemann, Wolffried Stucky (Hrsg.): Datenbanksysteme in Business, Technologie und Web

P-66 Jörg M. Haake, Ulrike Lucke, Djamshid Tavangarian (Hrsg.): DeLFI 2005: 3. deutsche e-Learning Fachtagung Informatik

P-67 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 1)

P-68 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 2)

P-69 Robert Hirschfeld, Ryszard Kowalcyk, Andreas Polze, Matthias Weske (Hrsg.): NODe 2005, GSEM 2005

P-70 Klaus Turowski, Johannes-Maria Zaha (Hrsg.): Component-oriented Enterprise Application (COAE 2005)

P-71 Andrew Torda, Stefan Kurz, Matthias Rarey (Hrsg.): German Conference on Bioinformatics 2005

P-72 Klaus P. Jantke, Klaus-Peter Fähnrich, Wolfgang S. Wittig (Hrsg.): Marktplatz Internet: Von e-Learning bis e-Payment

P-73 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): “Heute schon das Morgen sehen“

P-74 Christopher Wolf, Stefan Lucks, Po-Wah Yau (Hrsg.): WEWoRC 2005 – Western European Workshop on Research in Cryptology

P-75 Jörg Desel, Ulrich Frank (Hrsg.): Enterprise Modelling and Information Systems Architecture

P-76 Thomas Kirste, Birgitta König-Riess, Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Informationssysteme – Potentiale, Hindernisse, Einsatz

P-77 Jana Dittmann (Hrsg.): SICHERHEIT 2006

P-78 K.-O. Wenkel, P. Wagner, M. Morgenstern, K. Luzi, P. Eisermann (Hrsg.): Land- und Ernährungswirtschaft im Wandel

P-79 Bettina Biel, Matthias Book, Volker Gruhn (Hrsg.): Softwareengineering 2006

P-80 Mareike Schoop, Christian Huemer, Michael Rebstock, Martin Bichler (Hrsg.): Service-Oriented Electronic Commerce

P-81 Wolfgang Karl, Jürgen Becker, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle (Hrsg.): ARCS´06

P-82 Heinrich C. Mayr, Ruth Breu (Hrsg.): Modellierung 2006

P-83 Daniel Huson, Oliver Kohlbacher, Andrei Lupas, Kay Nieselt and Andreas Zell (eds.): German Conference on Bioinformatics

P-84 Dimitris Karagiannis, Heinrich C. Mayr, (Hrsg.): Information Systems Technology and its Applications

P-85 Witold Abramowicz, Heinrich C. Mayr, (Hrsg.): Business Information Systems

P-86 Robert Krimmer (Ed.): Electronic Voting 2006

P-87 Max Mühlhäuser, Guido Rößling, Ralf Steinmetz (Hrsg.): DELFI 2006: 4. e-Learning Fachtagung Informatik

P-88 Robert Hirschfeld, Andreas Polze, Ryszard Kowalczyk (Hrsg.): NODe 2006, GSEM 2006

P-90 Joachim Schelp, Robert Winter, Ulrich Frank, Bodo Rieger, Klaus Turowski (Hrsg.): Integration, Informationslogistik und Architektur

P-91 Henrik Stormer, Andreas Meier, Michael Schumacher (Eds.): European Conference on eHealth 2006

P-92 Fernand Feltz, Benoît Otjacques, Andreas Oberweis, Nicolas Poussing (Eds.): AIM 2006

P-93 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 1

P-94 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 2

P-95 Matthias Weske, Markus Nüttgens (Eds.): EMISA 2005: Methoden, Konzepte und Technologien für die Entwicklung von dienstbasierten Informationssystemen

P-96 Saartje Brockmans, Jürgen Jung, York Sure (Eds.): Meta-Modelling and Ontologies

P-97 Oliver Göbel, Dirk Schadt, Sandra Frings, Hardo Hase, Detlef Günther, Jens Nedon (Eds.): IT-Incident Mangament & IT-Forensics – IMF 2006

P-98 Hans Brandt-Pook, Werner Simonsmeier und Thorsten Spitta (Hrsg.): Beratung in der Softwareentwicklung – Modelle, Methoden, Best Practices

P-99 Andreas Schwill, Carsten Schulte, Marco Thomas (Hrsg.): Didaktik der Informatik

P-100 Peter Forbrig, Günter Siegel, Markus Schneider (Hrsg.): HDI 2006: Hochschuldidaktik der Informatik

P-101 Stefan Böttinger, Ludwig Theuvsen, Susanne Rank, Marlies Morgenstern (Hrsg.): Agrarinformatik im Spannungsfeld zwischen Regionalisierung und globalen Wertschöpfungsketten

P-102 Otto Spaniol (Eds.): Mobile Services and Personalized Environments

P-103 Alfons Kemper, Harald Schöning, Thomas Rose, Matthias Jarke, Thomas Seidl, Christoph Quix, Christoph Brochhaus (Hrsg.): Datenbanksysteme in Business, Technologie und Web (BTW 2007)

P-104 Birgitta König-Ries, Franz Lehner, Rainer Malaka, Can Türker (Hrsg.) MMS 2007: Mobilität und mobile Informationssysteme

P-105 Wolf-Gideon Bleek, Jörg Raasch, Heinz Züllighoven (Hrsg.) Software Engineering 2007

P-106 Wolf-Gideon Bleek, Henning Schwentner, Heinz Züllighoven (Hrsg.) Software Engineering 2007 – Beiträge zu den Workshops

P-107 Heinrich C. Mayr, Dimitris Karagiannis (eds.) Information Systems Technology and its Applications

P-108 Arslan Brömme, Christoph Busch, Detlef Hühnlein (eds.) BIOSIG 2007: Biometrics and Electronic Signatures

P-109 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 1

P-110 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 2

P-111 Christian Eibl, Johannes Magenheim, Sigrid Schubert, Martin Wessner (Hrsg.) DeLFI 2007: 5. e-Learning Fachtagung Informatik

P-112 Sigrid Schubert (Hrsg.) Didaktik der Informatik in Theorie und Praxis

P-113 Sören Auer, Christian Bizer, Claudia Müller, Anna V. Zhdanova (Eds.) The Social Semantic Web 2007 Proceedings of the 1st Conference on Social Semantic Web (CSSW)

P-114 Sandra Frings, Oliver Göbel, Detlef Günther, Hardo G. Hase, Jens Nedon, Dirk Schadt, Arslan Brömme (Eds.) IMF2007 IT-incident management & IT-forensics Proceedings of the 3rd International Conference on IT-Incident Management & IT-Forensics

P-115 Claudia Falter, Alexander Schliep, Joachim Selbig, Martin Vingron and Dirk Walther (Eds.) German conference on bioinformatics GCB 2007

P-116 Witold Abramowicz, Leszek Maciszek (Eds.) Business Process and Services Computing 1st International Working Conference on Business Process and Services Computing BPSC 2007

P-117 Ryszard Kowalczyk (Ed.) Grid service engineering and manegement The 4th International Conference on Grid Service Engineering and Management GSEM 2007

P-118 Andreas Hein, Wilfried Thoben, Hans-Jürgen Appelrath, Peter Jensch (Eds.) European Conference on ehealth 2007

P-119 Manfred Reichert, Stefan Strecker, Klaus Turowski (Eds.) Enterprise Modelling and Information Systems Architectures Concepts and Applications

P-120 Adam Pawlak, Kurt Sandkuhl, Wojciech Cholewa, Leandro Soares Indrusiak (Eds.) Coordination of Collaborative Engineering - State of the Art and Future Challenges

P-121 Korbinian Herrmann, Bernd Bruegge (Hrsg.) Software Engineering 2008 Fachtagung des GI-Fachbereichs Softwaretechnik

P-122 Walid Maalej, Bernd Bruegge (Hrsg.) Software Engineering 2008 - Workshopband Fachtagung des GI-Fachbereichs Softwaretechnik

P-123 Michael H. Breitner, Martin Breunig, Elgar Fleisch, Ley Pousttchi, Klaus Turowski (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Technologien, Prozesse, Marktfähigkeit
Proceedings zur 3. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2008)

P-124 Wolfgang E. Nagel, Rolf Hoffmann, Andreas Koch (Eds.)
9th Workshop on Parallel Systems and Algorithms (PASA)
Workshop of the GI/ITG Speciel Interest Groups PARS and PARVA

P-125 Rolf A.E. Müller, Hans-H. Sundermeier, Ludwig Theuvsen, Stephanie Schütze, Marlies Morgenstern (Hrsg.)
Unternehmens-IT:
Führungsinstrument oder Verwaltungsbürde
Referate der 28. GIL Jahrestagung

P-126 Rainer Gimnich, Uwe Kaiser, Jochen Quante, Andreas Winter (Hrsg.)
10th Workshop Software Reengineering (WSR 2008)

P-127 Thomas Kühne, Wolfgang Reisig, Friedrich Steimann (Hrsg.)
Modellierung 2008

P-128 Ammar Alkassar, Jörg Siekmann (Hrsg.)
Sicherheit 2008
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 4. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
2.-4. April 2008
Saarbrücken, Germany

P-129 Wolfgang Hesse, Andreas Oberweis (Eds.)
Sigsand-Europe 2008
Proceedings of the Third AIS SIGSAND European Symposium on Analysis, Design, Use and Societal Impact of Information Systems

P-130 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
1. DFN-Forum Kommunikations-technologien Beiträge der Fachtagung

P-131 Robert Krimmer, Rüdiger Grimm (Eds.)
3rd International Conference on Electronic Voting 2008
Co-organized by Council of Europe, Gesellschaft für Informatik and E-Voting. CC

P-132 Silke Seehusen, Ulrike Lucke, Stefan Fischer (Hrsg.)
DeLFI 2008:
Die 6. e-Learning Fachtagung Informatik

P-133 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik
Band 1

P-134 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik
Band 2

P-135 Torsten Brinda, Michael Fothe, Peter Hubwieser, Kirsten Schlüter (Hrsg.)
Didaktik der Informatik –
Aktuelle Forschungsergebnisse

P-136 Andreas Beyer, Michael Schroeder (Eds.)
German Conference on Bioinformatics GCB 2008

P-137 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2008: Biometrics and Electronic Signatures

P-138 Barbara Dinter, Robert Winter, Peter Chamoni, Norbert Gronau, Klaus Turowski (Hrsg.)
Synergien durch Integration und Informationslogistik
Proceedings zur DW2008

P-139 Georg Herzwurm, Martin Mikusz (Hrsg.)
Industrialisierung des Software-Managements
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik

P-140 Oliver Göbel, Sandra Frings, Detlef Günther, Jens Nedon, Dirk Schadt (Eds.)
IMF 2008 - IT Incident Management & IT Forensics

P-141 Peter Loos, Markus Nüttgens, Klaus Turowski, Dirk Werth (Hrsg.)
Modellierung betrieblicher Informations-systeme (MobIS 2008)
Modellierung zwischen SOA und Compliance Management

P-142 R. Bill, P. Korduan, L. Theuvsen, M. Morgenstern (Hrsg.)
Anforderungen an die Agrarinformatik durch Globalisierung und Klimaveränderung

P-143 Peter Liggesmeyer, Gregor Engels, Jürgen Münch, Jörg Dörr, Norman Riegel (Hrsg.)
Software Engineering 2009
Fachtagung des GI-Fachbereichs Softwaretechnik

P-144 Johann-Christoph Freytag, Thomas Ruf, Wolfgang Lehner, Gottfried Vossen (Hrsg.)
Datenbanksysteme in Business, Technologie und Web (BTW)

P-145 Knut Hinkelmann, Holger Wache (Eds.)
WM2009: 5th Conference on Professional Knowledge Management

P-146 Markus Bick, Martin Breunig, Hagen Höpfner (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Entwicklung, Implementierung und Anwendung
4. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2009)

P-147 Witold Abramowicz, Leszek Maciaszek, Ryszard Kowalczyk, Andreas Speck (Eds.)
Business Process, Services Computing and Intelligent Service Management
BPSC 2009 · ISM 2009 · YRW-MBP 2009

P-148 Christian Erfurth, Gerald Eichler, Volkmar Schau (Eds.)
9th International Conference on Innovative Internet Community Systems
I^{2}CS 2009

P-149 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
2. DFN-Forum Kommunikationstechnologien
Beiträge der Fachtagung

P-150 Jürgen Münch, Peter Liggesmeyer (Hrsg.)
Software Engineering
2009 - Workshopband

P-151 Armin Heinzl, Peter Dadam, Stefan Kirn, Peter Lockemann (Eds.)
PRIMIUM
Process Innovation for Enterprise Software

P-152 Jan Mendling, Stefanie Rinderle-Ma, Werner Esswein (Eds.)
Enterprise Modelling and Information Systems Architectures
Proceedings of the 3rd Int'l Workshop EMISA 2009

P-153 Andreas Schwill, Nicolas Apostolopoulos (Hrsg.)
Lernen im Digitalen Zeitalter
DeLFI 2009 – Die 7. E-Learning Fachtagung Informatik

P-154 Stefan Fischer, Erik Maehle Rüdiger Reischuk (Hrsg.)
INFORMATIK 2009
Im Focus das Leben

P-155 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2009:
Biometrics and Electronic Signatures
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures

P-156 Bernhard Koerber (Hrsg.)
Zukunft braucht Herkunft
25 Jahre »INFOS – Informatik und Schule«

P-157 Ivo Grosse, Steffen Neumann, Stefan Posch, Falk Schreiber, Peter Stadler (Eds.)
German Conference on Bioinformatics 2009

P-158 W. Claupein, L. Theuvsen, A. Kämpf, M. Morgenstern (Hrsg.)
Precision Agriculture Reloaded – Informationsgestützte Landwirtschaft

P-159 Gregor Engels, Markus Luckey, Wilhelm Schäfer (Hrsg.)
Software Engineering 2010

P-160 Gregor Engels, Markus Luckey, Alexander Pretschner, Ralf Reussner (Hrsg.)
Software Engineering 2010 – Workshopband
(inkl. Doktorandensymposium)

P-161 Gregor Engels, Dimitris Karagiannis Heinrich C. Mayr (Hrsg.)
Modellierung 2010

P-162 Maria A. Wimmer, Uwe Brinkhoff, Siegfried Kaiser, Dagmar Lück-Schneider, Erich Schweighofer, Andreas Wiebe (Hrsg.)
Vernetzte IT für einen effektiven Staat
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2010

P-163 Markus Bick, Stefan Eulgem, Elgar Fleisch, J. Felix Hampe, Birgitta König-Ries, Franz Lehner, Key Pousttchi, Kai Rannenberg (Hrsg.)
Mobile und Ubiquitäre Informationssysteme
Technologien, Anwendungen und Dienste zur Unterstützung von mobiler Kollaboration

P-164 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2010: Biometrics and Electronic Signatures Proceedings of the Special Interest Group on Biometrics and Electronic Signatures

P-165 Gerald Eichler, Peter Kropf, Ulrike Lechner, Phayung Meesad, Herwig Unger (Eds.)
10th International Conference on Innovative Internet Community Systems (I^2CS) – Jubilee Edition 2010 –

P-166 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
3. DFN-Forum Kommunikationstechnologien
Beiträge der Fachtagung

P-167 Robert Krimmer, Rüdiger Grimm (Eds.)
4th International Conference on Electronic Voting 2010
co-organized by the Council of Europe, Gesellschaft für Informatik and E-Voting.CC

P-168 Ira Diethelm, Christina Dörge, Claudia Hildebrandt, Carsten Schulte (Hrsg.)
Didaktik der Informatik
Möglichkeiten empirischer Forschungsmethoden und Perspektiven der Fachdidaktik

P-169 Michael Kerres, Nadine Ojstersek Ulrik Schroeder, Ulrich Hoppe (Hrsg.)
DeLFI 2010 - 8. Tagung der Fachgruppe E-Learning der Gesellschaft für Informatik e.V.

P-170 Felix C. Freiling (Hrsg.)
Sicherheit 2010
Sicherheit, Schutz und Zuverlässigkeit

P-171 Werner Esswein, Klaus Turowski, Martin Juhrisch (Hrsg.)
Modellierung betrieblicher Informationssysteme (MobIS 2010)
Modellgestütztes Management

P-172 Stefan Klink, Agnes Koschmider Marco Mevius, Andreas Oberweis (Hrsg.)
EMISA 2010
Einflussfaktoren auf die Entwicklung flexibler, integrierter Informationssysteme
Beiträge des Workshops der GI-Fachgruppe EMISA (Entwicklungsmethoden für Informationssysteme und deren Anwendung)

P-173 Dietmar Schomburg, Andreas Grote (Eds.)
German Conference on Bioinformatics 2010

P-174 Arslan Brömme, Torsten Eymann, Detlef Hühnlein, Heiko Roßnagel, Paul Schmücker (Hrsg.)
perspeGKtive 2010
Workshop „Innovative und sichere Informationstechnologie für das Gesundheitswesen von morgen“

P-175 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 1

P-176 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 2

P-177 Witold Abramowicz, Rainer Alt, Klaus-Peter Fähnrich, Bogdan Franczyk, Leszek A. Maciaszek (Eds.)
INFORMATIK 2010
Business Process and Service Science – Proceedings of ISSS and BPSC

P-178 Wolfram Pietsch, Benedikt Krams (Hrsg.)
Vom Projekt zum Produkt
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschafts-informatik (WI-MAW), Aachen, 2010

P-179 Stefan Gruner, Bernhard Rumpe (Eds.)
FM+AM`2010
Second International Workshop on Formal Methods and Agile Methods

P-180 Theo Härder, Wolfgang Lehner, Bernhard Mitschang, Harald Schöning, Holger Schwarz (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW)
14. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme“ (DBIS)

P-181 Michael Clasen, Otto Schätzel, Brigitte Theuvsen (Hrsg.)
Qualität und Effizienz durch informationsgestützte Landwirtschaft, Fokus: Moderne Weinwirtschaft

P-182 Ronald Maier (Hrsg.)
6th Conference on Professional Knowledge Management
From Knowledge to Action

P-183 Ralf Reussner, Matthias Grund, Andreas Oberweis, Walter Tichy (Hrsg.)
Software Engineering 2011
Fachtagung des GI-Fachbereichs Softwaretechnik

P-184 Ralf Reussner, Alexander Pretschner, Stefan Jähnichen (Hrsg.)
Software Engineering 2011
Workshopband
(inkl. Doktorandensymposium)

P-185 Hagen Höpfner, Günther Specht, Thomas Ritz, Christian Bunse (Hrsg.)
MMS 2011: Mobile und ubiquitäre Informationssysteme Proceedings zur 6. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2011)

P-186 Gerald Eichler, Axel Küpper, Volkmar Schau, Hacène Fouchal, Herwig Unger (Eds.)
11th International Conference on Innovative Internet Community Systems (I^2CS)

P-187 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
4. DFN-Forum Kommunikations-technologien, Beiträge der Fachtagung 20. Juni bis 21. Juni 2011 Bonn

P-188 Holger Rohland, Andrea Kienle, Steffen Friedrich (Hrsg.)
DeLFI 2011 – Die 9. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. 5.–8. September 2011, Dresden

P-189 Thomas, Marco (Hrsg.)
Informatik in Bildung und Beruf
INFOS 2011
14. GI-Fachtagung Informatik und Schule

P-190 Markus Nüttgens, Oliver Thomas, Barbara Weber (Eds.)
Enterprise Modelling and Information Systems Architectures (EMISA 2011)

P-191 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2011
International Conference of the Biometrics Special Interest Group

P-192 Hans-Ulrich Heiß, Peter Pepper, Holger Schlingloff, Jörg Schneider (Hrsg.)
INFORMATIK 2011
Informatik schafft Communities

P-193 Wolfgang Lehner, Gunther Piller (Hrsg.)
IMDM 2011

P-194 M. Clasen, G. Fröhlich, H. Bernhardt, K. Hildebrand, B. Theuvsen (Hrsg.)
Informationstechnologie für eine nachhaltige Landbewirtschaftung
Fokus Forstwirtschaft

P-195 Neeraj Suri, Michael Waidner (Hrsg.)
Sicherheit 2012
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 6. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)

P-196 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2012
Proceedings of the 11th International Conference of the Biometrics Special Interest Group

P-197 Jörn von Lucke, Christian P. Geiger, Siegfried Kaiser, Erich Schweighofer, Maria A. Wimmer (Hrsg.)
Auf dem Weg zu einer offenen, smarten und vernetzten Verwaltungskultur
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2012

P-198 Stefan Jähnichen, Axel Küpper, Sahin Albayrak (Hrsg.)
Software Engineering 2012
Fachtagung des GI-Fachbereichs Softwaretechnik

P-199 Stefan Jähnichen, Bernhard Rumpe, Holger Schlingloff (Hrsg.)
Software Engineering 2012
Workshopband

P-200 Gero Mühl, Jan Richling, Andreas Herkersdorf (Hrsg.)
ARCS 2012 Workshops

P-201 Elmar J. Sinz Andy Schürr (Hrsg.)
Modellierung 2012

P-202 Andrea Back, Markus Bick, Martin Breunig, Key Pousttchi, Frédéric Thiesse (Hrsg.)
MMS 2012:Mobile und Ubiquitäre Informationssysteme

P-203 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
5. DFN-Forum Kommunikations-technologien
Beiträge der Fachtagung

P-204 Gerald Eichler, Leendert W. M. Wienhofen, Anders Kofod-Petersen, Herwig Unger (Eds.)
12th International Conference on Innovative Internet Community Systems (I2CS 2012)

P-205 Manuel J. Kripp, Melanie Volkamer, Rüdiger Grimm (Eds.)
5th International Conference on Electronic Voting 2012 (EVOTE2012)
Co-organized by the Council of Europe, Gesellschaft für Informatik and E-Voting.CC

P-206 Stefanie Rinderle-Ma, Mathias Weske (Hrsg.)
EMISA 2012
Der Mensch im Zentrum der Modellierung

P-207 Jörg Desel, Jörg M. Haake, Christian Spannagel (Hrsg.)
DeLFI 2012: Die 10. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V.
24.–26. September 2012

P-208 Ursula Goltz, Marcus Magnor, Hans-Jürgen Appelrath, Herbert Matthies, Wolf-Tilo Balke, Lars Wolf (Hrsg.) INFORMATIK 2012

P-209 Hans Brandt-Pook, André Fleer, Thorsten Spitta, Malte Wattenberg (Hrsg.) Nachhaltiges Software Management

P-210 Erhard Plödereder, Peter Dencker, Herbert Klenk, Hubert B. Keller, Silke Spitzer (Hrsg.) Automotive – Safety & Security 2012 Sicherheit und Zuverlässigkeit für automobile Informationstechnik

P-211 M. Clasen, K. C. Kersebaum, A. Meyer-Aurich, B. Theuvsen (Hrsg.) Massendatenmanagement in der Agrar- und Ernährungswirtschaft Erhebung - Verarbeitung - Nutzung Referate der 33. GIL-Jahrestagung 20. – 21. Februar 2013, Potsdam

P-212 Arslan Brömme, Christoph Busch (Eds.) BIOSIG 2013 Proceedings of the 12th International Conference of the Biometrics Special Interest Group 04.–06. September 2013 Darmstadt, Germany

P-213 Stefan Kowalewski, Bernhard Rumpe (Hrsg.) Software Engineering 2013 Fachtagung des GI-Fachbereichs Softwaretechnik

P-214 Volker Markl, Gunter Saake, Kai-Uwe Sattler, Gregor Hackenbroich, Bernhard Mit schang, Theo Härder, Veit Köppen (Hrsg.) Datenbanksysteme für Business, Technologie und Web (BTW) 2013 13. – 15. März 2013, Magdeburg

P-215 Stefan Wagner, Horst Lichter (Hrsg.) Software Engineering 2013 Workshopband (inkl. Doktorandensymposium) 26. Februar – 1. März 2013, Aachen

P-216 Gunter Saake, Andreas Henrich, Wolfgang Lehner, Thomas Neumann, Veit Köppen (Hrsg.) Datenbanksysteme für Business, Technologie und Web (BTW) 2013 – Workshopband 11. – 12. März 2013, Magdeburg

P-217 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.) 6. DFN-Forum Kommunikations- technologien Beiträge der Fachtagung 03.–04. Juni 2013, Erlangen

P-218 Andreas Breiter, Christoph Rensing (Hrsg.) DeLFI 2013: Die 11 e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI) 8. – 11. September 2013, Bremen

P-219 Norbert Breier, Peer Stechert, Thomas Wilke (Hrsg.) Informatik erweitert Horizonte INFOS 2013 15. GI-Fachtagung Informatik und Schule 26. – 28. September 2013

P-220 Matthias Horbach (Hrsg.) INFORMATIK 2013 Informatik angepasst an Mensch, Organisation und Umwelt 16. – 20. September 2013, Koblenz

P-221 Maria A. Wimmer, Marijn Janssen, Ann Macintosh, Hans Jochen Scholl, Efthimios Tambouris (Eds.) Electronic Government and Electronic Participation Joint Proceedings of Ongoing Research of IFIP EGOV and IFIP ePart 2013 16. – 19. September 2013, Koblenz

P-222 Reinhard Jung, Manfred Reichert (Eds.) Enterprise Modelling and Information Systems Architectures (EMISA 2013) St. Gallen, Switzerland September 5. – 6. 2013

P-223 Detlef Hühnlein, Heiko Roßnagel (Hrsg.) Open Identity Summit 2013 10. – 11. September 2013 Kloster Banz, Germany

P-224 Eckhart Hanser, Martin Mikusz, Masud Fazal-Baqaie (Hrsg.) Vorgehensmodelle 2013 Vorgehensmodelle – Anspruch und Wirklichkeit 20. Tagung der Fachgruppe Vorgehensmodelle im Fachgebiet Wirtschaftsinformatik (WI-VM) der Gesellschaft für Informatik e.V. Lörrach, 2013

P-225 Hans-Georg Fill, Dimitris Karagiannis, Ulrich Reimer (Hrsg.) Modellierung 2014 19. – 21. März 2014, Wien

P-226 M. Clasen, M. Hamer, S. Lehnert, B. Petersen, B. Theuvsen (Hrsg.) IT-Standards in der Agrar- und Ernährungswirtschaft Fokus: Risiko- und Krisenmanagement Referate der 34. GIL-Jahrestagung 24. – 25. Februar 2014, Bonn

P-227 Wilhelm Hasselbring,
Nils Christian Ehmke (Hrsg.)
Software Engineering 2014
Fachtagung des GI-Fachbereichs Softwaretechnik
25. – 28. Februar 2014
Kiel, Deutschland

P-228 Stefan Katzenbeisser, Volkmar Lotz, Edgar Weippl (Hrsg.)
Sicherheit 2014
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 7. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
19. – 21. März 2014, Wien

P-230 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2014
Proceedings of the 13th International Conference of the Biometrics Special Interest Group
10. – 12. September 2014 in Darmstadt, Germany

P-231 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
7. DFN-Forum Kommunikationstechnologien
16. – 17. Juni 2014
Fulda

P-232 E. Plödereder, L. Grunske, E. Schneider, D. Ull (Hrsg.)
INFORMATIK 2014
Big Data – Komplexität meistern
22. – 26. September 2014
Stuttgart

P-233 Stephan Trahasch, Rolf Plötzner, Gerhard Schneider, Claudia Gayer, Daniel Sassiat, Nicole Wöhrle (Hrsg.)
DeLFI 2014 – Die 12. e-Learning Fachtagung Informatik
der Gesellschaft für Informatik e.V.
15. – 17. September 2014
Freiburg

P-234 Fernand Feltz, Bela Mutschler, Benoît Otjacques (Eds.)
Enterprise Modelling and Information Systems Architectures
(EMISA 2014)
Luxembourg, September 25-26, 2014

P-235 Robert Giegerich,
Ralf Hofestädt,
Tim W. Nattkemper (Eds.)
German Conference on Bioinformatics 2014
September 28 – October 1
Bielefeld, Germany

P-236 Martin Engstler, Eckhart Hanser, Martin Mikusz, Georg Herzwurm (Hrsg.)
Projektmanagement und Vorgehensmodelle 2014
Soziale Aspekte und Standardisierung
Gemeinsame Tagung der Fachgruppen Projektmanagement (WI-PM) und Vorgehensmodelle (WI-VM) im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V., Stuttgart 2014

P-237 Detlef Hühnlein, Heiko Roßnagel (Hrsg.)
Open Identity Summit 2014
4.–6. November 2014
Stuttgart, Germany

P-238 Arno Ruckelshausen, Hans-Peter Schwarz, Brigitte Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Referate der 35. GIL-Jahrestagung
23. – 24. Februar 2015, Geisenheim

P-239 Uwe Aßmann, Birgit Demuth, Thorsten Spitta, Georg Püschel, Ronny Kaiser (Hrsg.)
Software Engineering & Management 2015
17.-20. März 2015, Dresden

P-240 Herbert Klenk, Hubert B. Keller, Erhard Plödereder, Peter Dencker (Hrsg.)
Automotive – Safety & Security 2015
Sicherheit und Zuverlässigkeit für automobile Informationstechnik
21.–22. April 2015, Stuttgart

P-241 Thomas Seidl, Norbert Ritter, Harald Schöning, Kai-Uwe Sattler, Theo Härder, Steffen Friedrich, Wolfram Wingerath (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2015)
04. – 06. März 2015, Hamburg

P-242 Norbert Ritter, Andreas Henrich, Wolfgang Lehner, Andreas Thor, Steffen Friedrich, Wolfram Wingerath (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2015) – Workshopband
02. – 03. März 2015, Hamburg

P-243 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
8. DFN-Forum Kommunikationstechnologien
06.–09. Juni 2015, Lübeck

P-244 Alfred Zimmermann,
Alexander Rossmann (Eds.)
Digital Enterprise Computing
(DEC 2015)
Böblingen, Germany June 25-26, 2015

P-246 Douglas W. Cunningham, Petra Hofstedt,
Klaus Meer, Ingo Schmitt (Hrsg.)
INFORMATIK 2015
28.9.-2.10. 2015, Cottbus

The titles can be purchased at:

Köllen Druck + Verlag GmbH
Ernst-Robert-Curtius-Str. 14 · D-53117 Bonn
Fax: +49 (0)228/9898222
E-Mail: druckverlag@koellen.de

Gesellschaft für Informatik e.V. (GI)

publishes this series in order to make available to a broad public recent findings in informatics (i.e. computer science and information systems), to document conferences that are organized in cooperation with GI and to publish the annual GI Award dissertation.

Broken down into
• seminars
• proceedings
• dissertations
• thematics
current topics are dealt with from the vantage point of research and development, teaching and further training in theory and practice. The Editorial Committee uses an intensive review process in order to ensure high quality contributions.

The volumes are published in German or English.

Information: http://www.gi.de/service/publikationen/lni/

ISSN 1617-5468
ISBN 978-3-88579-640-4

"INFORMATIK 2015" was the 45th Annual Meeting of the Gesellschaft für Informatik and took place at the Brandenburgische Technische Universität Cottbus-Senftenberg in Cottbus. The guiding theme of the meeting was the interplay between interdisciplinary methods relating the fields of Computer Science, Energy and Environment.
This volume contains contributions from the workshop program, the student and doctoral symposium, and the tutorials.